Spring 5-12-2021

Explorations of Thermus thermophilus HB8 Transcription Factors TTHA1437 and TTHA1719 using the Combinatorial Selection Method REPSA

Sarah Claire Sneed

Explorations of *Thermus thermophilus* HB8 Transcription Factors TTHA1437 and TTHA1719 using the Combinatorial Selection Method REPSA

by

Sarah Claire Sneed

B.S. in Biochemistry

Kennesaw State University, 2019

Submitted in Partial Fulfillment of the Requirements

For the Degree of Master of Science in the

Department of Chemistry and Biochemistry

Kennesaw State University

2021

DocuSigned by:
Michael Van Dyke
FFF164E8A9194D8...
Committee Chair

DocuSigned by:
Chris Dockery
C1295244895A414...
Graduate Program Coordinator

DocuSigned by:
Carol Christensen
0F81EBA0E2804BC...
Committee Member

DocuSigned by:
Mark Mitchell
68C4A21FAB064E5...
Department Chair

DocuSigned by:
5122FD0B6FC3427...
Committee Member

DocuSigned by:
Kojo Mensa-Wilmot
39226A8692FB45B...
College Dean

ACKNOWLEDGEMENTS

I would like to sincerely thank Dr. Michael Van Dyke for his assistance, advice, and invaluable input throughout these research projects; his passion for scientific knowledge is contagious. I would also like to thank Dr. Tsai-Tien Tseng and Dr. Carol Chrestensen for serving as members of my research committee for your time and giving me guidance and input on these projects. To the faculty and staff in the Department of Chemistry and Biochemistry, the College of Science and Mathematics, and the Graduate College at Kennesaw State University, thank you for your continued support. Thank you to the National Science Foundation for financially supporting these projects (Grant MCB 1714778).

I would like to thank my peers in the MSCB program and my fellow Van Dyke Lab members for your scientific input and encouragement. Moreover, I would like to thank my family, specifically Marylou Uland, Nancy Sneed, and Geoff Edwards, for allowing me this opportunity to further my education and for their continued support prior to and throughout this journey. Finally, I would like to thank anyone who takes the time to read this document in the hopes it will better their scientific knowledge.

ABSTRACT

In all organisms, genes exist that encode for regulatory proteins, called transcription factors (TFs), that can activate or repress transcription of specific genes depending on their biological function. *Thermus thermophilus* HB8 is speculated to contain 2,245 genes, of which 70 are postulated to be transcription factors. However, but for very few, little is known regarding the genes they regulate and their biological functions. The novel combinatorial method Restriction Endonuclease Protection, Selection, and Amplification (REPSA) has successfully identified and validated consensus binding sequences in *T. thermophilus* HB8 with previously studied TFs. Here REPSA was explored as a technique and method for characterizing TTHA1437 and TTHA1719 TFs in *T. thermophilus* HB8. TTHA1437 REPSA results showed a promising DNA selection, but the results were not reproducible. Contrarily, TTHA1719 REPSA results did not show any selected DNAs, but during REPSA selections, a rare DNA species, the asterisk species, was observed. REPA results for TTHA1437 showed nonspecific binding, and TTHA1719 results showed no validation. EMSA also exhibited no DNA-ligand complex formation for both TFs. From the literature, a proposed potential consensus DNA for TTHA1719 proved promising with initial REPAs, and possible homologous DNA-binding consensus sequences for TTHA1437 and *E. coli* CRP were explored. Unfortunately, these were ultimately unsuccessful. These results demonstrate that REPSA is not a viable method to characterize all TFs and prompted us to explore why REPSA does not work for these *T. thermophilus* HB8 TFs and how it could be

improved. One possible improvement could entail more extensive preliminary DNA-binding assays with potential DNA-binding site(s) before embarking on REPSA selections.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

BLI	Biolayer interferometry
cAMP	Cyclic adenosine monophosphate
CAP	Catabolite activator protein
CRP	cAMP receptor protein
DBD	DNA-binding domain
EMSA	Electrophoretic mobility shift assay
FNR	Fumarate and nitrate reductase regulatory protein
HGP	Human genome project
HTH	Helix-turn-helix
IISRE	Type IIS restriction endonuclease
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Lysogeny broth
NCBI	National center for biotechnology information
NGS	Next generation sequencing
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
REPA	Restriction endonuclease protection assay
REPSA	Restriction endonuclease protection, selection, and amplification
RNAP	Ribonucleic acid polymerase

SOB	Super optimal broth
TF	Transcription factor

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Chapter 1.1 Bacterial Transcription Factors

The vast amount of sequence data presently available on whole genomes (eukaryotic and prokaryotic) is staggering since the Human Genome Project (HGP) was completed in 2003, the price of sequencing technologies has significantly decreased thus allowing for affordable in-house Next Generation Sequencing (NGS).[1] Although many genomics, transcriptomics, and metabolomics data have been obtained since the completion of the HGP, the biological functions of many genes are still unknown and therefore are unable to be assigned. Thus, scientists turn to one of the simpler organisms in life, prokaryotes.[2] Despite the fact that eukaryotic organisms are genetically more diverse than prokaryotic organisms, there are in fact prokaryotes which are model organisms for understanding the biological functionsof genes, as well as gene regulation.[3,4] Some well-known model prokaryotes include *Escherichia coli* (*E. coli*), *Pseudomonas aeruginosa (P. aeruginosa),* and *Thermus thermophilus (T. thermophilus).*[5]

An extremely important cellular process throughout all organisms of life, transcription, converts the information encoded in the genome (DNA) to messenger RNA (mRNA), transfer RNA (tRNA), or ribosomal RNA (rRNA)[6]. The central dogma of molecular biology describes the flow of genetic information, DNA→RNA→protein. Without gene expression, cells would be unable to express proteins and therefore unable to perform even basic cellular functions. The transcription process is highly regulated, in part by proteins called transcription factors (TFs).[7] Without control by TFs there would be unnecessary transcripts being transcribed and necessary transcripts not being

transcribed, both outcomes are detrimental. TFs are an organism's key to gene regulation; either activating or repressing genes of interest during the initiation of transcription.[8] Transcription factors regulate gene expression by interacting with promoter and/or enhancer elements upstream of the transcriptional startsite. When they interact with these regulatory elements the gene of interest can be repressed or activated.[8,9] Transcriptional repressors block RNA Polymerase (RNAP) holoenzyme from binding promoter DNA, and therefore repressing transcription. In contrast, transcriptional activators recruit RNAP holoenzyme which then increases the frequency of transcription.[10,11] Although many TFs have been identified and characterized in *E. coli* and *P. aeruginosa*, there is a lack of knowledge in the extreme thermophile, *T. thermophilus.*

Chapter 1.2 Types of Transcriptional Regulators

A classic and extremely well characterized model for gene regulation is the *lac* operon in *Escherichia coli* which is involved in lactose metabolism.[12] The *lac* operon contains a single promoter region and consists of three structural genes denoted *lacZ, lacY,* and *lacA* which encode for β-galactosidase, galactoside permease, and galactoside acetyltransferase respectively. Although these genes are regulated in parallel, galactoside acetyltransferase does not play a major role in lactose metabolism.[12] Upstream of the structural genes, there exists a catabolite activator protein (CAP) binding site, a promoter, and an operator region which are involved in the regulation of the *lac* operon; therefore, transcription can be activated or repressed.

Transcription is initiated at the promoter site which is located just upstream of the structural gene *lacZ*. In the event that cellular concentrations of glucose are high and that lactose is unavailable, a repressor protein will bind to the operator region and the *lac* genes will not be expressed.[12,13] Contrarily, positive transcription control will occur if the cellular concentration of glucose is low and lactose is high. When glucose levels are low, levels of cAMP are high and readily bind with the catabolite activator protein (CAP).[13] Once the CAP-cAMP complex is formed, the transcriptional activator binds DNA within the promoter region and allows robust expression of the *lac* genes. The positive and negative regulation of the *lac* operon is extremely important for cells to adequately respond to environmental changes and stressors.[14,15]

CRPs, also known as catabolite activator proteins (CAPs), tend to be homologous throughout many organisms due to their similar amino acid residues, structural similarity, and helix-turn-helix (HTH) motifs.[16] The HTH motif, two α-helices joined by a short

loop of amino acids, is extremely important for DNA-binding proteins, especially transcription factors.[17] Many HTH proteins bind to DNA in a dimer fashion and recognize palindromic sequences as their DNA-binding domain (DBD).[17,18] Similarly, other DNA-binding proteins (e.g. restriction endonucleases), recognize palindromic DNA sequences, which is important for their cleavage specificity. Although there are proteins that recognize non-palindromic sequences, the majority of regulatory proteins prefer palindromic sequences such as the CRP global regulator in *E. coli*.[19]

In *Escherichia coli* there exists a set of transcription factors which are involved in the regulation of multiple genes when a cellular stress-response is activated. Interestingly, one of these TFs is the *E. coli* cAMP receptor protein (CRP); this CRP is a global regulator within *E. coli* due to the sheer number of genes this transcription factor regulates.[19,21] One unique property of *E. coli* CRP is its ability to act as an activator, repressor, coactivator and corepressor. This unique property, along with the different classes of promoters it recognizes, permits CRP to regulate a tremendous amount of *Escherichia coli's* genome.[21] Although it only directly regulates the transcription of ~300 of the ~4500 genes (depending on the strain), the *E. coli* CRP indirectly regulates gene expression of about half the genome.[22]

Global regulators and local regulators are defined by the number and the function of the genes they regulate. Global TFs regulate many operons with different metabolic functions, while local TFs regulate single gene pathways or operons which belong to the same functional classification. Interestingly, global regulators can interact with local regulators to form a co-regulation mechanism which can in turn lead to a feedback loop which can widely influence gene expression[22]. Although a vast amount of information is

known about the *Escherichia coli* CRP global gene regulator, there is very little knowledge available about the homologous *Thermus thermophilus* CRP, TTHA1437.

Chapter 1.3 Identification and Characterization of TFs in *Thermus thermophilus* HB8

In the past, the Van Dyke laboratory has used REPSA to identify transcriptional regulators in the extreme thermophile and model organism *Thermus thermophilus* HB8. This gram-negative, rod shaped, aerobic bacterium was first isolated from a Japanese hot spring in 1968 and can grow in temperatures ranging from 50°C to 82°C, with its optimal growth temperature being between 65°C and 72°C.[23,24] Although there are two well-studied strains of *T. thermophilus*, HB8 and HB27, our focus is on the HB8 strain given that it is the model organism for the Structural-Biological Whole Cell Project at the RIKEN Institute in Japan.[25] The Whole Cell Project aims to understand all biological phenomena within *T. thermophilus* HB8 at an atomic level through the determination of three-dimensional structures for all of its proteins. To date, the Structural-Biological Whole Cell Project has purified 944 *T. thermophilus* HB8 proteins and 682 of these have been crystallized.[25]

The genome of the HB8 strain consists of a ~1.85 Mbp circular chromosome (TTHA), a mega plasmid of ~0.26 Mbp (TTHB), and a mini plasmid of ~9.3 kbp (TTHC).[25] The difference between the genome of HB8 and HB27 is that HB8 has the addition of the mini plasmid, TTHC, which gives the HB8 strain a unique advantage as a facultative anaerobe versus an obligate anaerobe (HB27). The transcription factors of interest for this study are located on the circular chromosome and denoted TTHA1437 and TTHA1719. TTHA1437 is 23,819 Daltons (Da) in mass and is suspected to belong to the CRP family of activators in *T. thermophilus* HB8 including TTHA1437, TTHA1567 and TTHB099 due to their homology within the *E. coli* CRP global regulator.[20,26,27] TTHA1719 is a transcriptional repressor with a mass of 10,858 Da and is suspected to be

a copper homeostasis operon regulatory protein due to its homology with CsoR from *M. tuberculosis* and *B. subtilis*.[28]

Chapter 1.4 Significance of Restriction Endonuclease Protection, Selection, and Amplification for Identifying Preferred DNA-Binding Sites

Restriction Endonuclease Protection, Selection, and Amplification (REPSA) is a novel combinatorial method developed by the Van Dyke laboratory. This is an *in vitro* PCR-based technique that utilizes a pool of randomized DNA sequences to select for TF-DNA interactions. When a type IIS restriction endonuclease (IISRE) is introduced, any uncomplexed DNAs will be cleaved and intact templates will be PCR amplified to further select for a DNA-binding sequence. One of the key components of the REPSA selection method is the DNA selection template which is derived from the ST2R24 or ST2R35 template precursors (Figure 1A and 1B). Both selection templates are composed of a 23-mer (ST2L) primer on the 5' end and a fluorescently red-labeled 25-mer (IRD7_ST2R) primer on the 3'; both primers flank the internal randomized region (either 24-mer or 35-mer, respectively). The IRD7_ST2R primer was specifically designed to contain the binding/recognition sites for two differenttype IIS restriction endonucleases: FokI (CATCC) and BpmI (CTCCAG).[29]

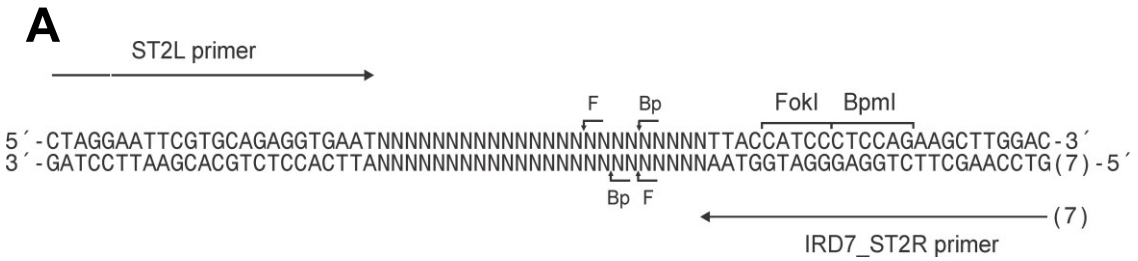

Figure 1A. REPSA selection templates ST2R24 and ST2R35. A depiction of the ST2R24 REPSA selection template and its components (ST2R35 not pictured). ST2L and IRD7_ST2R primers are denoted by horizontal arrows, (N) denotes random nucleotides within the randomized region (only difference between ST2R24 and ST2R35 are N=24 or N=35), and the brackets and small arrows denote the recognition domain and cleavage domain of IISREs FokI and BpmI.

Type IISREs are unique compared to other restriction endonucleases since they bind asymmetric dsDNA at a specific sequence (recognition domain) but cleave at a fixed distance away (cleavage domain); this distance varies for different IISREs. These types of restriction endonucleases are useful for many applications in molecular biology such as gene cloning and DNA fragmentation.[30] IISREs also allow theprobing of a randomized region to select for sequences that specifically bind a DNA- binding ligand (*e.g.*, transcription factor). This unique characteristic allows REPSA selections to work for a variety of different DNA-binding molecules (e.g. drugs, multiplex DNAs, proteins).[31]

The novel combinatorial method Restriction Endonuclease, Protection, Selection, and Amplification (REPSA) is a PCR-based method used to identify and characterize nucleic acid-protein interactions. REPSA experiments utilize the ST2R24 DNA selection template, IISREs, and a ligand of interest; in our case TTHA147.[29] A schematic for REPSA selection is shown in Figure 2. Here, the black regions represent the known flanking regions of the template while the red region represents the internal randomized region.[29,32,33] Once the ligand is introduced (green) it will bind to any preferred sequences that are present in the DNA pool. Due to the varying templates in the selection pool, the ligand will initially have very few templates to bind to. Thus, this selection method requires multiple rounds to identify a preferred DNA-binding site.

Once the ligand has been introduced to the selection templates, a IISRE is then introduced and any template with a ligand bound will not be cleaved, yet all unbound templates will be cleaved. An aliquot of the experimental reaction (DNA+IISRE+ligand) is then subjected to polymerase chain reaction (PCR), amplifying the selection templates

in 6, 9, and 12 cycles. The PCR reaction will not amplify any DNAs that have been cleaved, thus selecting for protected templates.[29]

During a round of REPSA, there are three reactions that run in parallel. The first being a control: DNA template, no IISRE, and no ligand (-/-). The second reaction is also a control, but a cleavage control: DNA template, IISRE, and no ligand (-/+). The third and most important reaction is the binding reaction: DNA template, IISRE, and TTHA1437. These reactions and the subsequent PCRs are visualized using 10% native PAGE and IR fluorescence to determine how much selection has occurred. The dsDNA concentration of the DNA pool is also quantified using a Qubit assay. Sequential rounds of REPSA are performed until ~50% cleavage inhibition is finally observed.

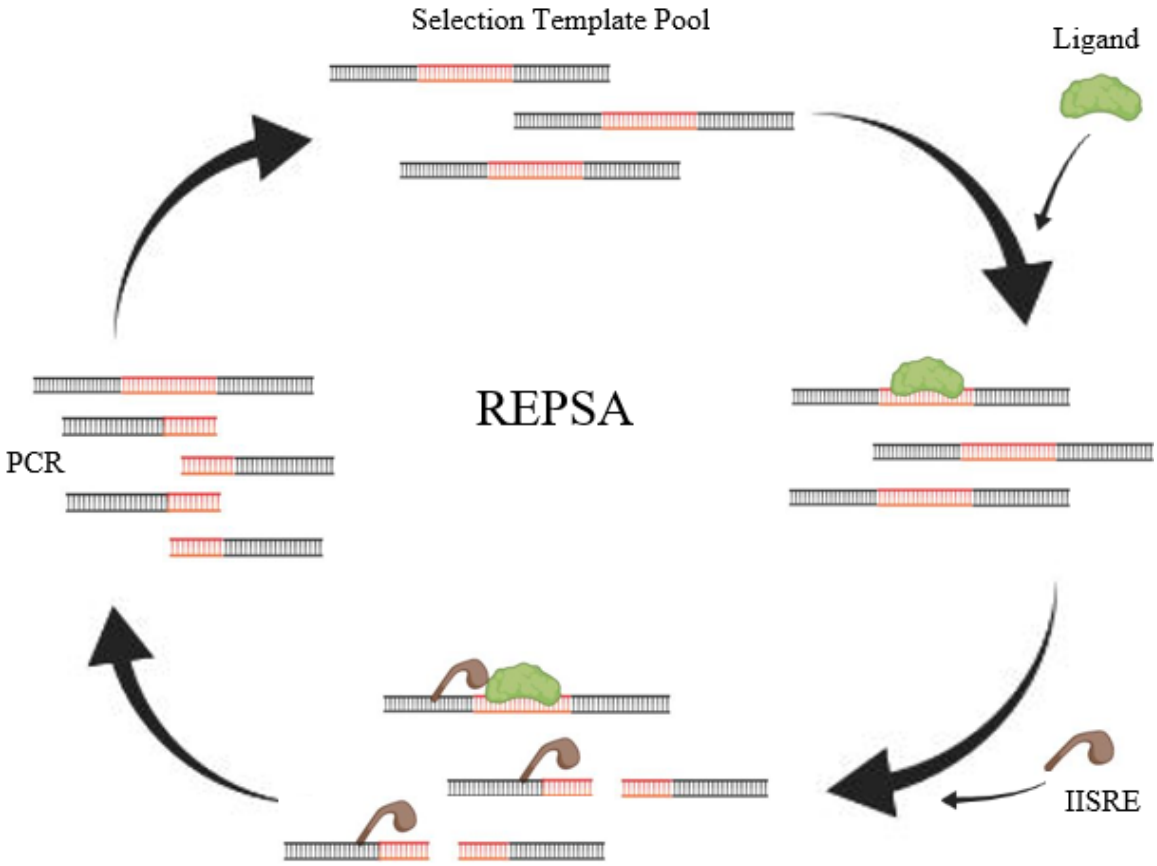

Figure 2. The REPSA method. A diagram depiction of restriction endonuclease protection, selection, and amplification (REPSA) and all the components included. The DNA selection template pool is symbolized by the black (primers) and red (24- or 35-mer randomized region). The ligand (TTHA1437 or TTHA1719) is shown in green, and the IISRE is shown in brown. The first step is the introduction of the ligand to the template pool at which time only a small percentage of DNAs will be bound to the ligand (complex formation step). The second step introduced a IISRE (either FokI or BpmI) which will cleave all DNA templates that are unbound by the ligand. After, this reaction will be subjected to PCR amplification but only the protected DNA species will amplify which further selects for the preferred DNA-binding sequence. Finally, the amplified DNAs will be used as the input for the next round of REPSA; subsequent rounds of REPSA will typically result in a selected DNA sequence which is ligand specific.

In Figure 3, rounds 1 and 7 of REPSA selections are shown for another *T. thermophilus* transcriptional regulator, SbtR, that has been previously studied.[32] Due to self-inhibitory cleavage by certain IISREs, the restriction endonuclease was switched from FokI to BpmI between rounds 4 and 5. In round 7, ~50% cleavage resistance due to sequence selection occurred, thus validating that this selection method is feasible, so the same experimental approach was to be used to identify and characterize TTHA1437 and TTHA1719.

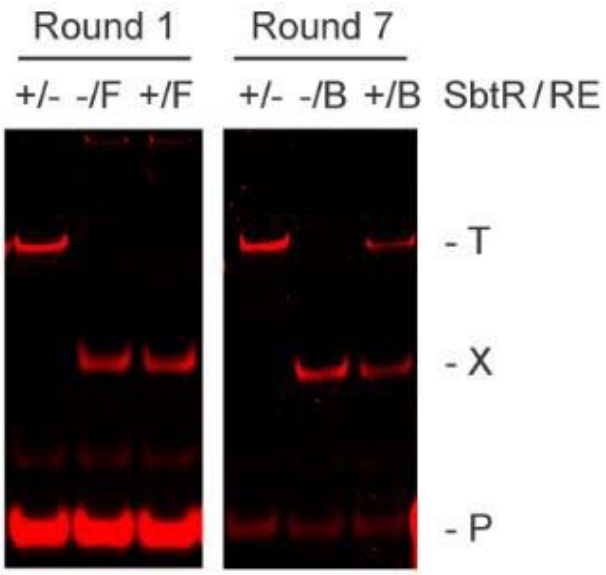

Figure 3. Expected REPSA results. Shown are IR fluorescence images of restriction endonuclease cleavage-protection assays for Round 1 and Round 7 of REPSA selections with SbtR protein. Reactions in lane 1 represents the DNA control (+/-), lane 2 being the cleavage control containing the DNA template and IISRE (-/F or -/B to represent FokI and BpmI IISREs respectively), and lane 3 is the experimental reaction containing DNA template, IISRE, and SbtR (+/F or +/B). Band designations: (T) intact ST2R24 selection template, (X) cleaved ST2R24 selection template, and (P) IRD7_ST2R primer. This image has been adapted from Figure 4 in Reference 32.

Once a pool of protected DNA templates has been selected, it is important to validate that the selection contains *bona fide* DNA-binding sites. It is important to validate the experimental results before sequencing due to the occurrence of non-specific binding artifacts. One validation experiment is Restriction Endonuclease Protection

Assay (REPA) which is similar to REPSA but, does not include the amplification step. An example is shown in Figure 4A for the previously studied SbtR.[32] REPA validations require a known DNA control (REPSAis) and the experimental REPSA DNA. Since TTHA1437 requires the effector molecule 3',5'-cyclic adenosine monophosphate (cAMP), REPA will be the preferred method for validating the selection. Another validation experiment is Electrophoretic Mobility Shift Assay (EMSA), also called a gel shift assay, which is an extremely common technique used to determine nucleic acid-protein interactions (Figure 4B).[33,34] Although EMSA is an important nucleic acid-ligand assay, if the protein of interest requires an effector molecule it is not the most effective method to validate a REPSA selection. Both REPA and EMSA methods should be viable to validate TTHA1719 REPSA-selected DNAs since it is not known to require an effector molecule to repress transcription.[28]

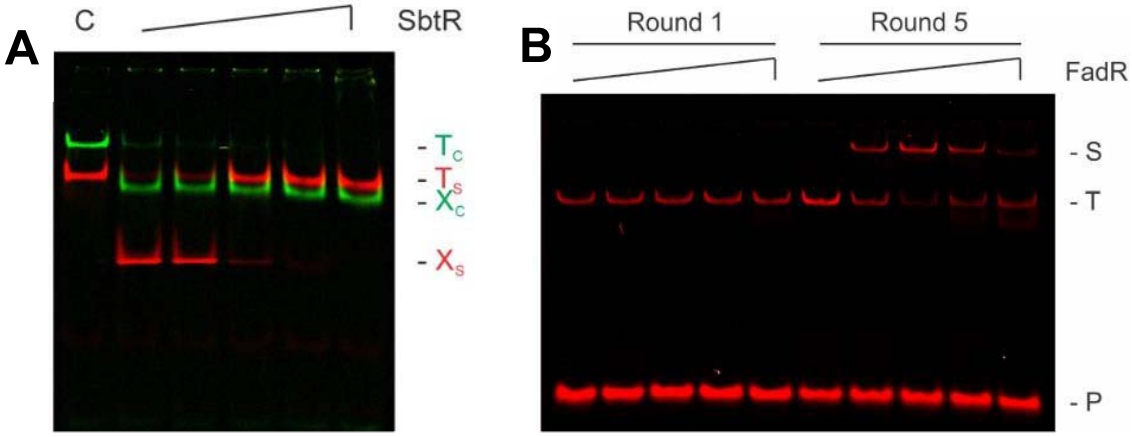

Figure 4. Traditional methods used to validate REPSA results. **(A)** Shown are LICOR images of IRD7-labeled SbtR consensus DNA (red) and IRD8-labeled REPSAis control DNA (green) subjected to BpmI cleavage following binding reactions in the presence of (left to right) 0, 1, 10, 100, 1000 nM dimer SbtR protein. (C) Uncleaved DNA control lane. (T) intact, uncleaved DNA, (X) cleaved DNA. **(B)** Shown are LICOR images of EMSAs containing pooled DNA from either Round 1 (left lanes) orRound 5 (right lanes) of REPSA selection and increasing concentrations of FadR protein (from left to right: 0, 0.6, 6, 60, or 600 nM FadR). The electrophoretic mobility of a single protein-DNA complex (S) as well as uncomplexed ST2R24 selection template (T) and IRD7_ST2R primer (P) are indicated at right of figure. These figures have been adapted from Figures 7 and Figure 2 from References 32 and 33 respectively.

Chapter 1.5 Investigation of Known and Proposed Consensus DNAs

For both *T. thermophilus* HB8 transcription factors in these studies, TTHA1437 and TTHA1719, there have been proposed potential consensus DNAs for their DNA-binding sites.[26,28] One of the problems that arises from these proposed consensus DNAs is that little, if any, binding assays were performed with their proposed consensus sequence(s). Although bioinformatic approaches can be extremely useful for determining possible ligand-nucleic acid analyses, it is important to test the findings of these bioinformatic studies to determine if it is in fact a true DNA-binding site. In order to validate these findings, it was important for us to investigate these proposed consensus sequences via different binding assays (REPA and EMSA). Furthermore, the DNA-binding consensus sequence for *E. coli* CRP was of interest in this study since TTHA1437 is a supposed homolog and belonging to the CRP/FNR group in *T. thermophilus*. The *E. coli* CRP consensus DNA-TTHA1437 interactions were to be investigated using EMSA and BLI.[22] The sequences of these known and proposed consensus DNAs can be found in Table 1.

Chapter 1.6 Hypothesis and Specific Aims

A reverse genetic approach can be used to identify and characterize *T. thermophilus* HB8 transcriptional regulators TTHA1437 and TTHA1719 and possible ascertain their biological functions.

1. Express and purify the proteins of interest, TTHA1437 and TTHA1719.
2. Use Restriction Endonuclease Protection, Selection, and Amplification (REPSA) to identify preferred DNA-binding sequences for TTHA1437 and TTHA1719.
3. Validate the REPSA selections via REPA, EMSA, and BLI.
4. Perform binding assays (REPA, EMSA, BLI) with proposed potential consensus DNAs and known consensus DNAs.
5. Analyses of results and bioinformatic studies, if REPSA-selected DNAs are validated, to determine the possible biological function of TTHA1437 and TTHA1719.

MATERIALS AND METHODS

Chapter 2.1 Oligonucleotides and Selection Template Preparation

The oligonucleotide precursors and primers used in these studies were synthesized and purified by Integrated DNA Technologies and are listed in Table 1. Two different selection templates, ST2R24 and ST2R35, were PCR amplified to create double-stranded DNA libraries for both templates. Each template was designed so the randomized cassette regions have an average nucleotide composition of 25% A, 25% T, 25% C, and 25% G at each position. Preparation of the ST2R24 template involved five 25 µL reactions in which each contained 1 ng single-stranded ST2R24 precursor DNAs, 1X Standard *Taq* Reaction Buffer (New England Biolabs, NEB), 560 nM ST2L primer, 560 nM IRD7_ST2R primer, 50 µM dNTPs, and 25 U *Taq* DNA Polymerase. These reactions were PCR amplified with the cycling conditions comprising of 5 cycles of 95°C for 30 sec, 60°C for 30 sec, and 68°C for 1 min; 1 cycle of 95°C for 30 sec, 60°C for 30 sec, and 68°C for 1.5 min; the final cycle of 95°C for 30 sec, 60°C for 30 sec, and 68°C for 2 min.

The ST2R35 template was prepared in a similar manner to ST2R24 although the reaction and cycling conditions differed slightly. Preparation of ST2R35 comprised of three 25 µL reactions with each containing 1 ng single-stranded ST2R35 precursor DNAs, 1X Standard *Taq* Reaction Buffer (NEB), 200 nM ST2Ls primer, 200 nM IRD7_ST2R primer, 200 µM dNTPs, and 12.5 U *Taq* DNA Polymerase. Cycling conditions involved 6 cycles of 95°C for 30 sec, 58°C for 1 min, and 68°C for 1.5 min. Although the construction of these selection templates differs, this allows for the

maximum amount of DNAs that contain a fully annealed random cassette region as well as increasing the diversity in the randomized region of the selection template.

The ST2_1437_jcon probe was prepared using a standard NEB PCR protocol for *Taq* DNA Polymerase which included 1 ng ST2_1437_jcon DNA, 1X Standard *Taq* Reaction Buffer, 200 nM ST2L primer, 200 nM IRD7_ST2R primer, 200 µM dNTPs, and 1.25 U *Taq* DNA Polymerase. The ST2_CRP_Ec probe was prepared in the same fashion as ST2_1437_jcon but 1 ng ST2_CRP_Ec was used instead. Cycling conditions consisted of 30 cycles of 90°C for 30 sec, 58°C for 30 sec, and 68°C for 1 min. The ST2_1719_jcon-60 and REPSAis probes were both prepared in the same fashion utilizing a fusion PCR protocol which involved two consecutive PCR amplifications. The first fusion PCR step mimics those of the ST2_1437_jcon preparation with 1 ng DNA template, ST2_1719_jcon-60 or REPSAis (ST2R_1719 and ST2Ls or trP1_ST2L and ST2R primers), but the samples were amplified for 6 cycles vs 30. After 6 rounds of amplification, 2 µL of the initial PCR was used to seed the next reaction containing 1X Standard *Taq* Reaction Buffer, 200 µM dNTPs, and 1.25 U *Taq* DNA Polymerase. For the ST2_1719_jcon-60 probe, 200 nM IRD7_ST2R and 200 nM ST2Ls primers were used and for REPSAis 200 nM IRD8_trP1_ST2L and 200 nM ST2R primers; both were amplified for 30 cycles as previously described. The REPSAis probe was used as an internal standard to determine if DNA-protein interactions are specific vs. nonspecific.

Table 1. Oligonucleotides used in these studies.

Name	Sequence	Length	Purif.	Use
ST2R24	CTAGGAATTCGTGCAGAGGTGAATNNNNNNNNNNNNNNNNNNNNNNNNTTACCATCCCTCCAGAAGCTTGGAC	73	PAGE	ST2R24 Template Precursor
ST2R35	CTAGGAATTCGTGCAGAGGTGAATNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTTACCATCCCTCCAGAAGCTTGGAC	84	PAGE	ST2R35 Template Precursor
ST2L	CTAGGAATTCGTGCAGAGGTGAAT	24	Desalt	PCR Left Primer
ST2Ls	CTAGGAATTCGTGCAGAGGTGA	22	Desalt	PCR Left Primer Short
ST2R	GTCCAAGCTTCTGGAGGGATGGTAA	25	Desalt	PCR Right Primer
IRD7_ST2R	/5IRD700/GTCCAAGCTTCTGGAGGGATGGTAA	25	HPLC	5'-IRDye700 PCR Primer
ST2R_1719	GTCCAAGCTTCTGGAGGGATGTACCCCA	28	Desalt	Fusion PCR Primer
trP1_ST2L	CCTCTCTATGGGCAGTCGGTGATCTAGGAATTCGTGCAGAGGTGA	45	PAGE	Fusion PCR Primer
ST2_1719_jcon-60	CTAGGAATTCGTGCAGAGGTGATACCCCACCCCACCTGGGGTGGGGTACATCCCTCCAGA	60	Desalt	TTHA1719 Japanese Consensus
ST2_1437_icon	CTAGGAATTCGTGCAGAGGTGATGTGCAAATTTCACACATCCCTCCAGAAGCTTGGAC	58	Desalt	TTHA1437 Japanese Consensus
IRD8_trP1_ST2L	/5IRD800/CCTCTCTATGGGCAGTCGGTGATCTAG	27	HPLC	5'-IRDye800-modified PCR Primer
Bio_ST2R	/5BiodT/GTCCAAGCTTCTGGAGGGATG	22	HPLC	5'-biotinylated PCR primer
ST2_CRP_Ec	GGAATTCGTGCAGAGGTGAAATGTGATCTAGATCACATTTCATCCCTCCAGAAGCTTGG	59	Desalt	*E. coli* CRP Consensus DNA
REPSAis	CTAGGAATTCGTGCAGAGGTGAATCGTCATAGAATTCGTTACCATCCCTCCAGAAGCTTGGAC	63	PAGE	Control DNA precursor

Chapter 2.2 Protein Preparation of TTHA1417 and TTHA1719

E. coli BL21 (DE3) competent cells were transformed with the pET-11a plasmid which contained the *TTHA1437* gene of interest. Once transformed, the cells were inoculated in Super Optimal Broth (SOB) media containing 100 μg/mL ampicillin and incubated at 37°C/250 rpm. The culture was induced with isopropyl β-D-1-thiogalactopyranoside (IPTG; 1 mM final concentration), incubated at 37°C/250 rpm for 4 hours to an OD_{600} of 0.651 and centrifuged to pellet the bacteria. The bacterial pellet was resuspended with 2X BEB (40 mM Tris-Cl [pH 7.5], 200 mM NaCl, 0.2 mM EDTA, 2 mM DTT, 1 mM PMSF) and stored at -20°C. Lysozyme was added to the thawed, resuspended sample to a final concentration of 0.25 μg/μL and incubated for 5 min at 0°C. The cells were then subjected to 5 rounds of sonication (3 W/cm^2, 10 sec on/10 sec off, 0°C) to lyse the cells and centrifuged (10 min, 13,000 rpm, 4°C) to pellet debris.

Following the steps above, the resulting supernatant was subjected to a simple heat treatment for 15 min at 70°C to purify TTHA1437. This purification method is extremely simple as the mesothermic *E. coli* proteins denature at 70°C while the thermophilic *T. thermophilus* proteins do not. Once purified, the sample was centrifuged (15 min, 13,000 rpm, 4°C). The supernatant was diluted in an equal volume of glycerol and rocked for 1 hour at 4°C to ensure a homogenous TTHA1417 protein stock for downstream experiments. Purified TTHA1417 was stored at -20°C. The purity of TTHA1437 was determined using 12% SDS-PAGE and its concentration was quantified utilizing densitometry with Coomassie and a BSA standard curve (0.5, 1.0, and 2.0 mg/mL).

TTHA1719 was prepared in a similar manner to TTHA1417, with the following alterations that allowed optimal protein expression. Due to the rare codon usage of TTHA1719, Rosetta^{TM} 2 (DE3) *E. coli* competent cells were transformed with the pTTHA1719 plasmid. 4 samples were then inoculated in 1 mL Lysogeny broth (LB) media in the presence of 100 µg/mL ampicillin and 35 µg/mL chloramphenicol and incubated at 37°C/250 rpm. After 1 hour, the two cloudiest cultures were transferred to 50 mL LB media in the presence of 100 µg/mL ampicillin and 35 µg/mL chloramphenicol, antibiotics required for the maintenance of plasmids. These samples were incubated at 37°C/250 rpm for 5 hours before 0.5 mL IPTG (1 mM final concentration) was added to the second sample (sample 1 was used as the uninduced control). Both samples were incubated at 37°C/250 rpm for 4 hours until the induced culture had an OD_{600} of 0.818 and the uninduced an OD_{600} of 1.23. These were centrifuged (15 min, 4,500 rpm, 4°C), the supernatant decanted, resuspended in 500 mL 2X BEB, and stored at -20°C until the purification process.

The samples were thawed and resuspended before 30 µL lysozyme (0.25 µg/µL final concentration) was added and incubated for 10 min at 0°C, mixing once halfway through. The cells were then subjected to 5 rounds of sonication (2.5 W/cm^2, 10 sec on/30 sec off, 0°C) and centrifuged (15 min, 13,000 rpm, 4°C). The supernatant was subjected to heat treatment at 70°C for 15 min, mixing halfway through, and centrifuged (15 min, 13,000 rpm, 4°C). An equal volume of glycerol was added to the supernatant and rocked (30 min, 4°C) to ensure proper mixing and was stored at -20°C for further use. The purity of TTHA1719 was determined similarly to TTHA1437 with the difference being an 18% SDS-PAGE containing 4 M urea. Due to the low molecular weight of TTHA1719, a traditional 12% SDS-PAGE would be ineffective. The

concentration of TTHA1437 was determined using a Bradford Protein Assay and the concentration of TTHA1719 was determined by Quantitative Densitometry with Coomassie; both of these assays use BSA as a standard. Although some *E. coli* and lysozyme proteins are still present, they have not interfered with or affected previously studied *T. thermophilus* TFs due to their denatured state and significant dilution of TTHA1437 and TTHA1719 in future experiments.

Chapter 2.3 REPSA Selection Method

REPSA selections for TTHA1437 were performed in 20 µL reactions with 4.515 ng (100 fmol) ST2R24 DNA template in 1X CutSmart® Buffer (NEB; 50 mM Potassium Acetate, 20 mM Tris-acetate, 10 mM Magnesium Acetate, 100 µg/mL BSA, pH 7.9 at 25°C) and 20 µM 3',5'-cAMP. The first round of REPSA utilized the initial ST2R24 DNA library, while successive REPSA sounds were seeded with 2 ng DNA from the previous round of REPSA. Each round of REPSA includes three reactions run in tandem: the DNA control, IISRE control, and the experimental reaction which contains the ligand of interest (-/-, -/IISRE, and +/IISRE). The DNA and IISRE controls contained 1-3 µL PDB (20 mM Tris HCl, 100 nM NaCl, 0.1 mM EDTA, 1 mM DTT, 100 µg/mL BSA, 0.1% Tween 20, pH 8.0 at 25°C) respectively, while the experimental reaction was incubated with 30.9 nM monomeric or 15.45 nM dimeric TTHA1437 protein diluted in PDB. Once assembled, these reactions were incubated at 55°C for 20 min to facilitate DNA-ligand interactions and then equilibrated at 37°C for 5 min. After equilibration, the IISRE control and experimental reactions were treated with 0.8 U FokI and 4 U BpmI REases for rounds 1-3 and 4-6 and incubated at 37°C for 5 min to cleave any unbound DNA species. The reactions were then placed on ice to cease endonuclease activity. Rounds of REPSA for TTHA1437 were also completed in a similar fashion, but without the addition of cAMP to determine if TTHA1437 does or does not require cAMP for DNA interactions to occur.

REPSA selections for TTHA1719 were performed in 10 µL reactions with 3.896 ng (75 fmoles) ST2R35 DNA selection template. The differences in REPSA selections for TTHA1719 include the addition of 1 mM DTT as a reducing agent and 41.6 nM

monomeric, 20.8 nM dimeric or 10.4 nM tetrameric TTHA1719 protein. Also, 0.8 U FokI and 4 U BpmI restriction endonucleases were used for rounds 1-2 and 3-11 respectively and incubated for 10 min to allow for maximum cleavage of unbound DNAs. All other aspects of these selections follow formerly outlined REPSA protocols.

After each round of REPSA, an amplification step is required before continuing on to another round of selections. Each amplification step utilized three 23 µL reactions which contained 1X NEB Standard *Taq* Reaction Buffer (NEB), 200 µM dNTPs, 200 nM ST2L or ST2Ls primer, 200 nM IRD7_ST2R primer, and 3.5 U *Taq* DNA Polymerase. 2 µL of the REPSA experimental reaction was added to each before PCR amplification. The reactions were then amplified for 6, 9, and 12 cycles utilizing the following cycling conditions: denaturation at 95°C for 30 sec, annealing at 58°C for 1 min, and elongation at 68°C for 1 min (1.5 min for ST2R35). After PCR amplification, 2 µL from each PCR reaction were combined with 2 µL 6X Orange Loading Dye (20% wt/vol dextrose, 0.9% wt/vol Orange G, 1% wt/vol SDS, and 66 mM EDTA).

REPSA reactions and the 6, 9, 12-cycle DNAs were run on a 10% wt/vol Native PAGE in 0.5X TBE for 10 min at 54V and then 55 min at 108V. Following gel electrophoresis, the results were imaged and visualized by LI-COR Odyssey Imager and the DNA concentrations of the amplified species were measured by Qubit 3 Fluorometer following the manufacturers protocol and reagents.

Chapter 2.4 Validation of REPSA Selections via Binding Assays

The selected DNA sequences from REPSA were subjected to Restriction Endonuclease Protection Assay (REPA) to validate the selections. REPA is similar to REPSA, but with the addition of the REPSAis control DNA and this assay does not include an amplification step. REPA 10 µL reactions were performed with 1X Cutsmart Buffer, 2 ng REPSA-selected DNAs (Round 6 for TTHA1437 and Round 11 for TTHA1719), 2 ng REPSAis control DNA, the respective transcription factor (TTHA1437 or TTHA1719), and the IISRE BpmI (4 U). REPA for TTHA1437 required 20 µM cAMP and TTHA1719 required 1 mM DTT. Reactions were incubated at 55°C for 20 minutes and then 37°C for 5 minutes before introducing 4 U BpmI and incubating for another 5-10 min at 37°C. 2 µL of each REPA reaction were mixed with 2 µL 6X Orange Loading Dye and run on a 10% wt/vol polyacrylamide (19:1 acryl:bis) gel at 54V for 10 min and 108V for 55 min before visualizing results by IR fluorescence.

Another validation method used was Electrophoretic Mobility Shift Assay (EMSA). EMSA 5 µL reactions were also performed with 2 ng Round 6 REPSA-selected DNA, PDB, 20 µM cAMP, and 1 µL TTHA1437 with ten-fold serial dilutions (0, 3.09, 30.9, 309, and 3,090 nM). The reactions were incubated at 55°C for 5 minutes to equilibrate prior to the addition of TTHA1437. After the addition of protein, the reactions were incubated at 55°C for 30 minutes to facilitate DNA-protein complex formation. 2 µL 6X Orange Loading Dye without SDS (20% wt/vol dextrose, 0.9% wt/vol Orange G, and 66 mM EDTA) was added to the EMSA samples prior to gel electrophoresis. The samples were run on a 0.5X TAE, 10% polyacrylamide gel and run in 0.5X TAE buffer

for 10 min at 54V and then 55 min at 108V. The results were visualized using the LI-COR Odyssey Imager as previously described.

Chapter 2.5 Binding Assays with Japanese Consensus Probes and *E. coli* Consensus Probe

Following REPSA selections for TTHA1437 and validation using REPA and EMSA, the Japanese consensus DNA for TTHA1437 was used in REPA and BLI. Similarly, the *E. coli* CRP consensus DNA was used in EMSA and BLI since TTHA1437 is a supposed orthologous TF to that of *E. coli* CRP. The REPA followed a similar protocol as previously described with 10 µL reactions containing 2 ng REPSAis control DNA, 2 ng ST2_1437_jcon DNA probe, 1X Cutsmart Buffer, and 20 µM cAMP. Reactions were incubated at 55°C for 20 min and then 37°C for 5 min to equilibrate reactions before the addition of 0.8 U FokI and a final incubation for 5 min at 37°C. Results were visualized by IR fluorescence as previously described. The EMSA done using the *E. coli* CRP consensus DNA was performed using the same protocol as previously outlined with the only difference being the DNA template (ST2_CRP_Ec vs REPSA selected DNAs) and dimeric TTHA1437 protein concentrations of 3.09, 9.27, 27.81, 83.43, and 250.29 nM.

Real-time binding kinetics were determined by biolayer interferometry (BLI) using biotinylated ST2_1437_jcon DNAs and TTHA1437. Biotinylated DNAs were prepared by PCR amplification in 50 µL reactions containing 1X Standard *Taq* Reaction Buffer, 50 µM dNTPs, 350 nM ST2L primer, 300 nM Bio_ST2R primer, 2 U *Taq* DNA Polymerase and 2 ng DNA template (ST2_1437_jcon or ST2_CRP_Ec consensus DNAs). Cycling conditions consisted of 30 cycles of 95°C for 30 sec, 58°C for 30 sec, and 68°C for 1 min. The BLI assays were performed with a FortéBio Octet^{QK} instrument in 96-well microplates using Streptavidin Biosensors (FortéBio) and biotinylated DNAs

prepared as previously described. Each BLI assay consisted of four lanes by four rows and each well contained 200 µL reactions which were buffered with BLI 100 Buffer supplemented with cAMP (20 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, 0.05% Tween 20, 20 µM cAMP, pH 7.7 at 25°C). The loading step consisted of 2 nM biotinylated DNAs, the background and dissociation steps consisted of 200 µL BLI 100 Buffer w/ cAMP, and the association step contained four concentrations of TTHA1437 (11, 33, 100, and 300 nM), respectively. Binding kinetics were determined using association and dissociation data and non-linear regression analyses in GraphPad Prism 8.

Unlike the binding assays for TTHA1437 using known DNA templates which were performed after REPSA selections, the binding assays for TTHA1719 and ST2_1719_jcon-60 were completed prior to REPSA selections in order to determine not only if binding occurs but also to determine the optimal protocol conditions to be used in REPSA selections. The first REPA was followed a similar protocol as previously described with 10 µL reactions containing 1X Cutmast Buffer, 2 ng REPSAis control DNA, 2 ng ST2_1719_jcon-60 test DNA, and ten-fold serial dilutions of TTHA1719. Reactions were incubated at 55°C for 20 min and then at 37°C for 5 min before the addition of 0.8U FokI and a final incubation at 37°C for 5 min. The results were visualized using IR fluorescence and were used as a baseline to compare other REPA experiments in which the reaction conditions were altered.

Because TTHA1719 contains a single cysteine residue, a reducing agent is needed to avoid disulfide bond formation. Another REPA was performed in the same fashion as the baseline REPA but with the addition of 2 mM DTT. Four other REPA experiments were performed to determine if altering the reaction parameters

would increase or decrease the binding activity of TTHA1719. Two of these REPAs consisted of altering the binding temperature from 55°C to 37°C and 65°C while in the other two REPAs the digestion time for FokI was changed from 5 min to 7 min and 9 minutes. All four of the REPA experiments with altered conditions were done under reducing conditions with 2 mM DTT and results were visualized by IR fluorescence on the LI-COR Odyssey Imager as previously described.

RESULTS

Chapter 3.1 Protein Expression, Purification, and Quantitation of TTHA1437 and TTHA1719

The TTHA1437 protein was expressed in *E. coli* BL21 (DE3) cells with a pET-11a plasmid containing the *TTHA1437* gene which yielded significant protein production. Once expressed, the protein was purified via heat treatment at 70°C and results were analyzed by 12% SDS-PAGE. Each step of the expression and purification can be visualized in Figure 5. TTHA1437 can be observed as a strong band with a molecular weight of about 24 kDa, which is consistent with the literature.[26] When comparing the soluble and purified phases, it is estimated that TTHA1437 is greater than 90% pure (Figure 5, lane 6). Although there are a few denatured *E. coli* proteins present (lane 6), but they are at such low concentrations that they should not affect any of our experimental results which has been previously observed for other *Thermus thermophilus* HB8 TFs studied in our laboratory.[32,33,35,36] The identities of these remaining soluble *E. coli* proteins have yet to be determined. TTHA1437 protein quantitation was done using a Bradford Protein Assay and its concentration was determined to be 30.9 μM (Figure 6).

Similarly, TTHA1719 was expressed in Rosetta 2 *E. coli* (DE3) cells with a pET-11a plasmid containing the *TTHA1719* gene; TTHA1719 was expressed in Rosetta 2 due to its rare codon usage. The results were analyzed by 18% SDS-PAGE with 4 M urea and the results can be visualized in Figure 7 with TTHA1719 as a strong band in lane 4. TTHA1719 was determined to be greater than 60% pure which is a lower purity than

normally observed for other *T. thermophilus* HB8 TFs that have been previously studied. The concentration of TTHA1719 was determined using Quantitative Densitometry with Coomassie and was calculated to be 41.6 μM (Figure 8).

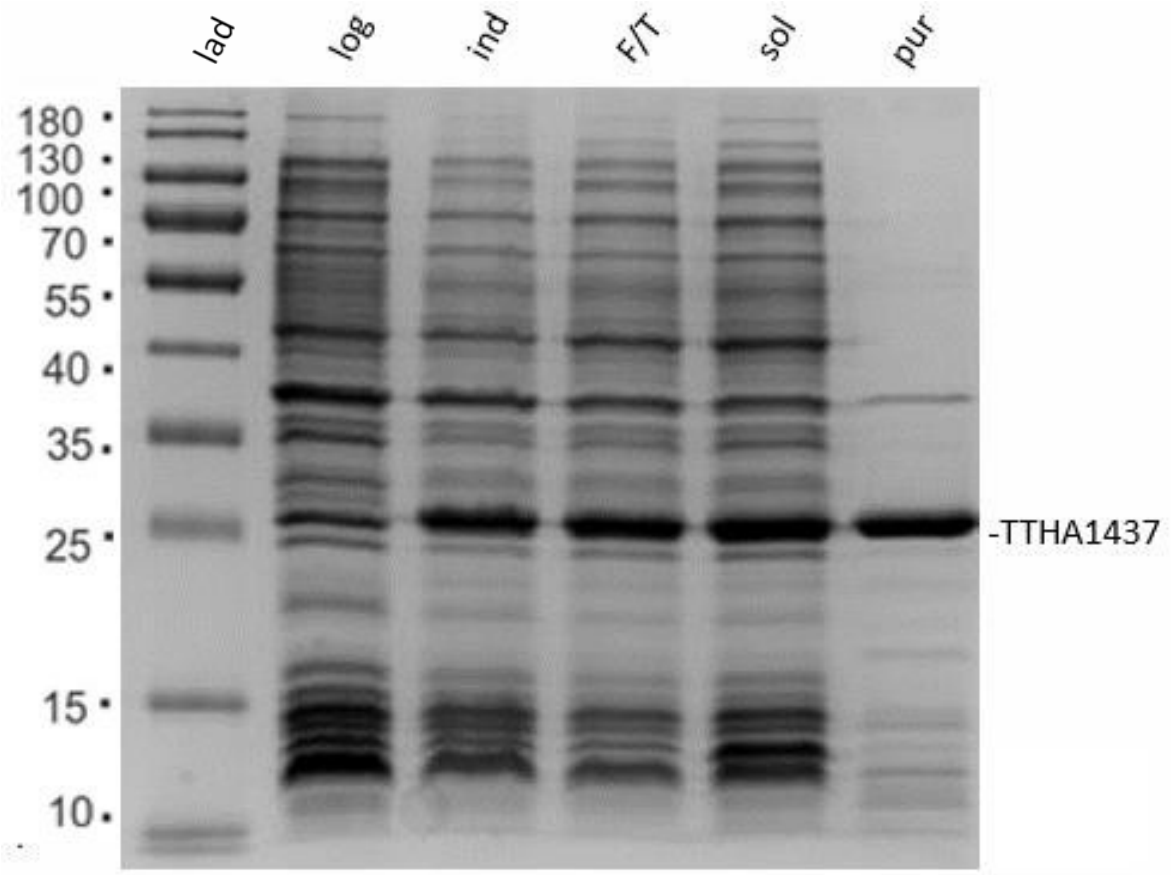

Figure 5. Expression and purification of TTHA1437 protein. Shown is a 12% SDS-PAGE gel stained with Coomassie Blue G-250 on which samples from protein expression and purification steps were analyzed. Lanes were loaded with samples of protein ladder (lad), logarithmic growth phase (log), induced and post-incubation (ind), sample after freeze and thaw cycle (F/T), soluble proteins following sonication (sol), and purified TTHA1437 following high temperature purification. Molecular weights of the protein ladder are indicated on the left of the figure and the protein band corresponding to TTHA1437 is indicated on the right.

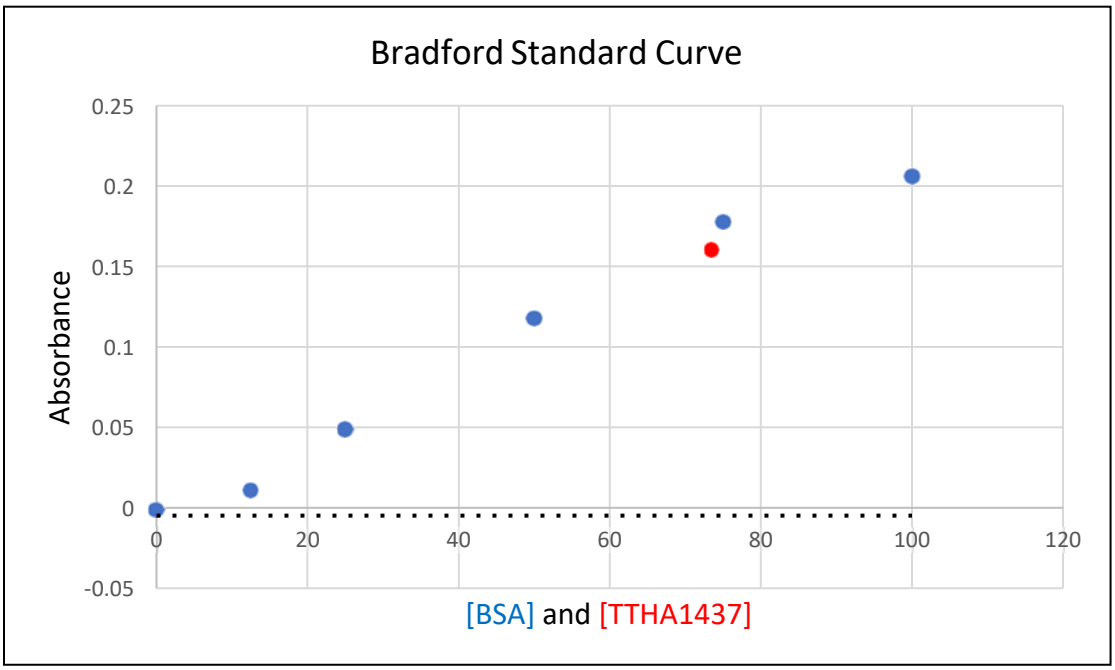

Figure 6. Quantitation of TTHA1437 using Bradford Protein Assay. Shown is the Bradford standard curve which was created by using bovine serum albumin (BSA) as the known standard. After the standard curve was established, the concentration of TTHA1437 was calculated using its molecular weight. The final concentration of TTHA1437 is estimated to be 30.9 µM.

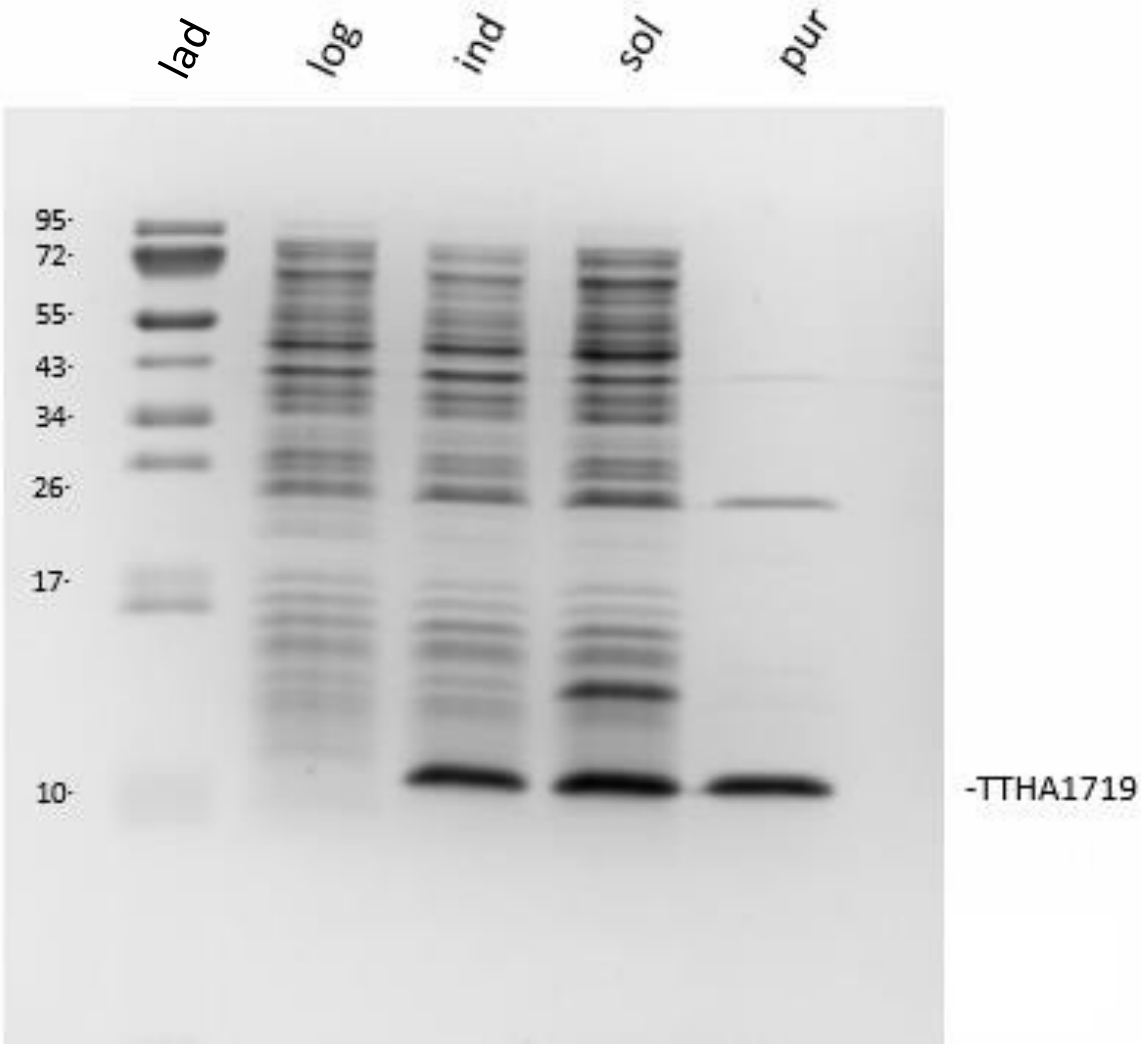

Figure 7. Expression and purification of TTHA1719 protein. Shown is an 18% SDS-PAGE w/ 4 M urea gel stained with Coomassie Blue G-250 on which samples from protein expression and purification steps were analyzed. Lanes were loaded with samples of protein ladder (lad), logarithmic growth phase (log), induced and post-incubation (ind), soluble proteins following sonication (sol), and purified TTHA1719 following high temperature purification. Molecular weights of the protein ladder are indicated on the left of the figure and the protein band corresponding to TTHA1719 is indicated on the right.

Figure 8. Quantitation of TTHA1719 using Quantitative Densitometry with Coomassie. This quantitation method also uses BSA as a standard to create a linear relationship, but the proteins are quantified by intensity rather than absorbance. Shown is the protein ladder (lad), BSA standards (left to right: 500 ng and 1000 ng protein), and TTHA1719. The final concentration of TTHA1719 was estimated to be 41.6 μM.

Chapter 3.2 REPSA Results for TTHA1437 and TTHA1719

REPSA was used to select the TTHA1437 and TTHA1719 binding sites from a pool of 100 fmoles or 75 fmoles DNA selection template molecules, respectively. The ST2R24 selection template was used in REPSA selections for TTHA1437, while the ST2R35 selection template was used for TTHA1719 REPSA selections. The ST2R24 template has previously been used to successfully identify TFs in other studies, but this is the first time that the ST2R35 template has been used in our laboratory for REPSA selections[32,33,35,36]. Here, TTHA1437 was subjected to two different REPSA selections in which one was run in the presence of 2 mM cAMP while the other was not. For TTHA1437 REPSA selections containing cAMP, six rounds were performed until the emergence of DNA resistant to IISRE cleavage (Figure 9, Round 6). In Round 6, the ST2R24 template DNA control (-/-) was uncut in the absence of BpmI and TTHA1437; the template DNA in the cleavage control (-/B) was completely cut with the presence of BpmI and the absence of TTA1437; the template DNA in the experimental reaction (+/B) was ~30% uncut in the presence of BpmI and TTHA1437 which represents the REPSA-selected sequences. Similarly, TTHA1437 REPSA selections were done without the addition of cAMP and were performed for nine rounds until IISRE cleavage resistance was observed in the cleavage control and experimental reaction (Figure 10, Round 9, lanes 2 and 3).

TTHA1719 REPSA selections were performed for ten rounds using the ST2R35 DNA template under reducing conditions with 1 mM DTT until ~10% of the template was uncut (Figure 11, Round 10, lane 3) and the internally primed asterisk species was greater than 40% (Figure 11, Round 10, lane 1). It is important to note that initial rounds

of REPSA utilized the IISRE FokI until FokI resistant DNAs emerged in the cleavage control at which point subsequent REPSA rounds used the IISRE BpmI to cleave the DNA selection templates. Typically, REPSA results are validated by REPA and EMSA before sequencing, but since the REPSA-selected sequences were not validated, the selections were not sequenced.

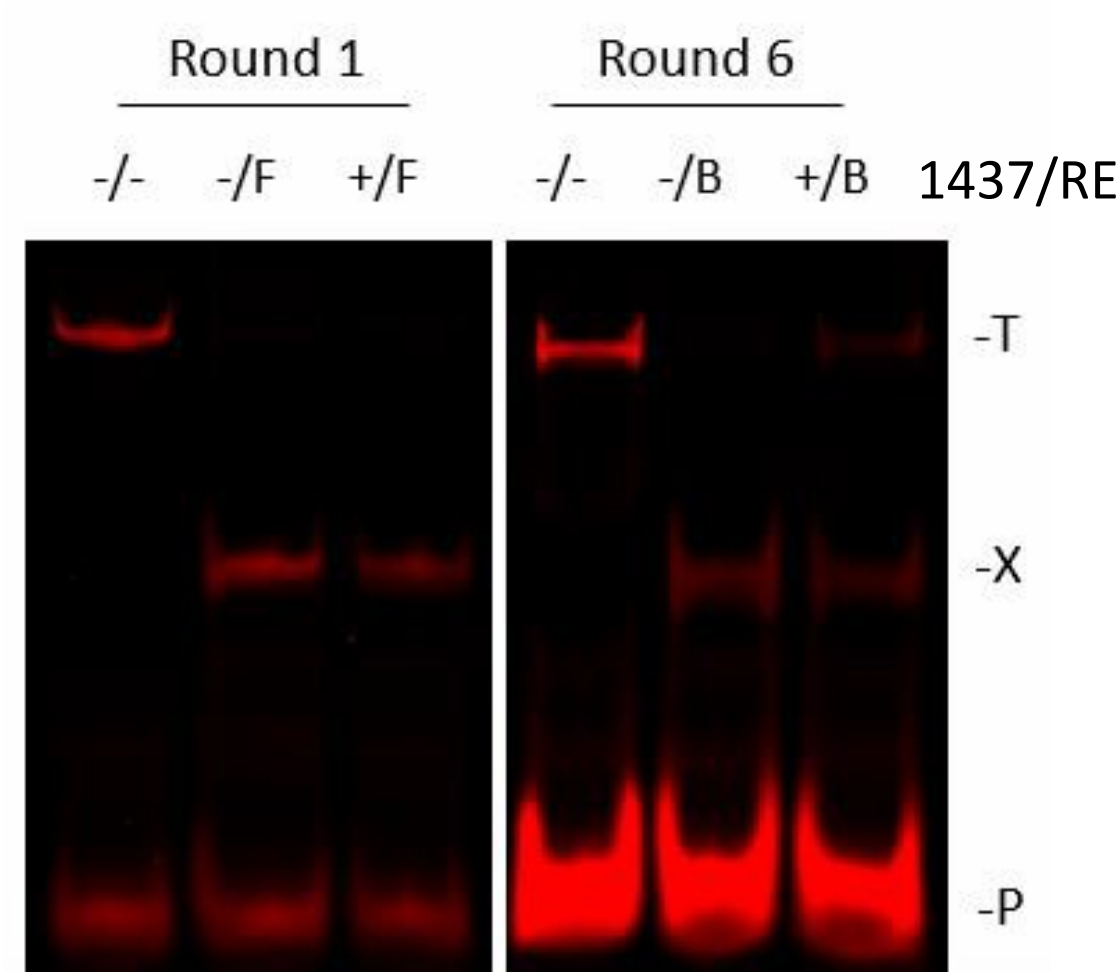

Figure 9. REPSA selection of TTHA1437-binding sequences with cAMP. Shown are IR fluorescence images of restriction endonuclease cleavage-protection assays for Round 1 and Round 6 of REPSA selections with 31 nM TTHA1437 protein. Reactions in lane 1 represents the DNA control (-/-), lane 2 being the cleavage control containing the DNA template and IISRE (-/F or -/B to represent FokI andBpmI IISREs respectively), and lane 3 is the experimental reaction containing DNA template, IISRE, and TTHA1437 (+/F or +/B). Band designations: (T) intact ST2R24 selection template, (X) cleaved ST2R24 selection template, and (P) IRD7_ST2R primer.

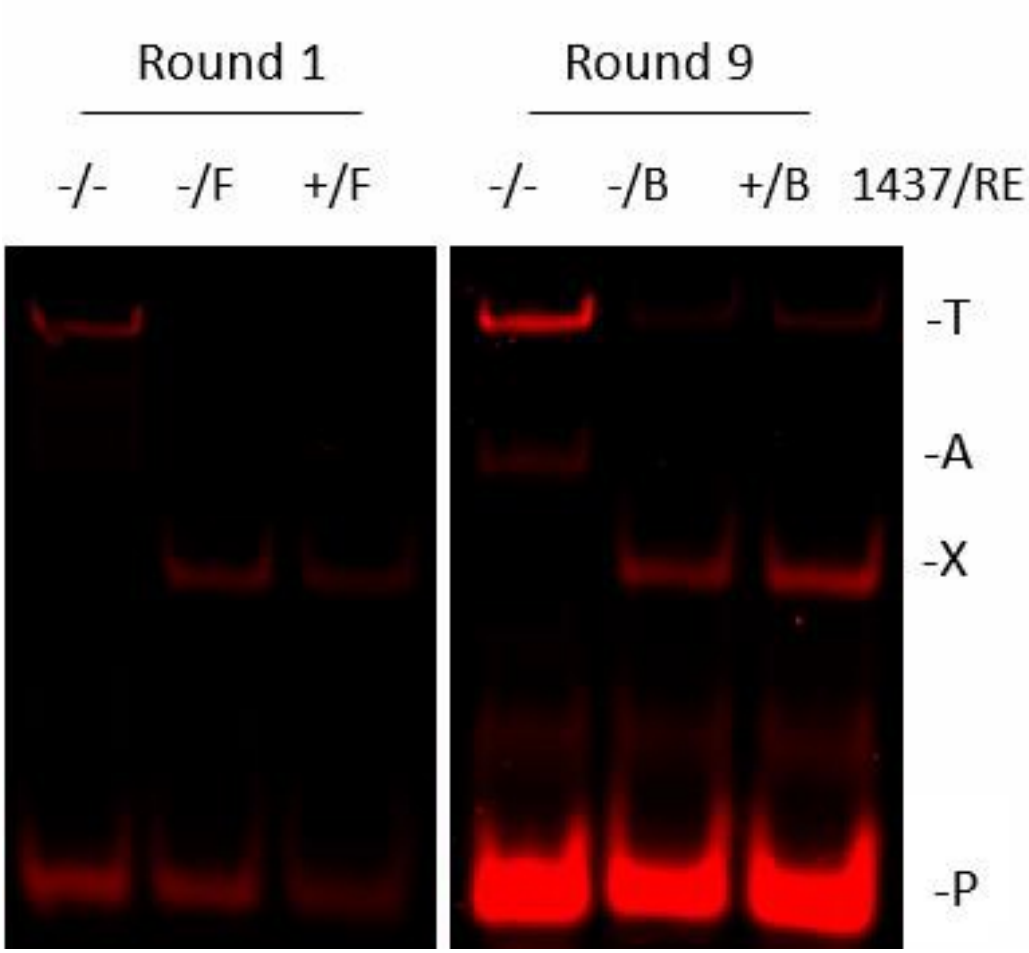

Figure 10. REPSA selection of TTHA1437-binding sequences without cAMP. Shown are IR fluorescence images of restriction endonuclease cleavage-protection assays for Round 1 and Round 9 of REPSA selections with 31 nM TTHA1437 protein. Reactions in lane 1 represents the DNA control (-/-), lane 2 being the cleavage control containing the DNA template and IISRE (-/F or -/B to represent FokI andBpmI IISREs respectively), and lane 3 is the experimental reaction containing DNA template, IISRE, and TTHA1437 (+/F or +/B). Band designations: (T) intact ST2R24 selection template, (A) internally primedasterisk species, (X) cleaved ST2R24 selection template, and (P) IRD7_ST2R primer.

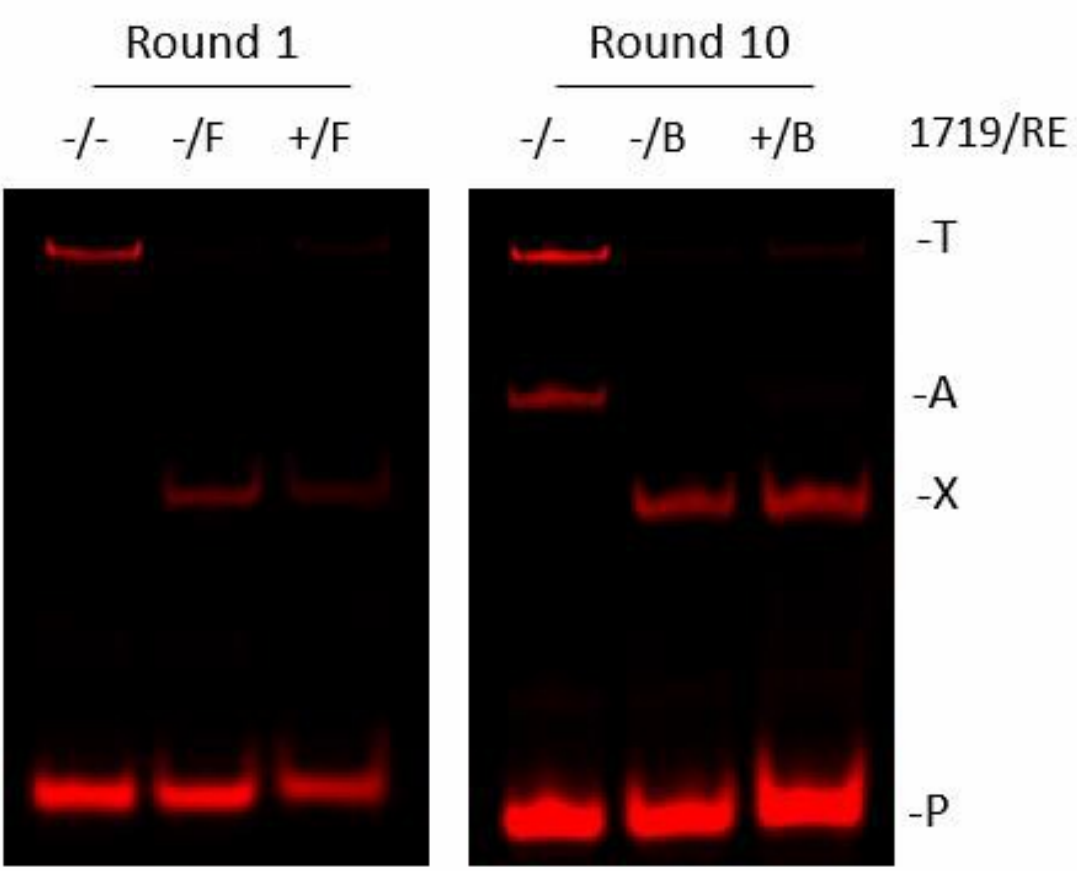

Figure 11. REPSA selection of TTHA1719-binding sequences. Shown are IR fluorescence images of restriction endonuclease cleavage-protection assays for Round 1 and Round 10 of REPSA selections with 42 nM TTHA1719 protein. Reactions in lane 1 represents the DNA control (-/-), lane 2 being the cleavage control containing the DNA template and IISRE (-/F or -/B to represent FokI and BpmI IISREs respectively), and lane 3 is the experimental reaction containing DNA template, IISRE, and TTHA1437 (+/F or +/B). Band designations: (T) intact ST2R35 selection template, (A) internally primed asterisk species, (X) cleaved ST2R35 selection template, and (P) IRD7_ST2R primer.

Chapter 3.3 Validation of REPSA Results via REPA and EMSA

Following REPSA selections for TTHA1437 and TTHA1719, the selected DNA species were attempted to be validated using both REPA and EMSA, but both were unsuccessful. REPA was performed to determinewhether specific or nonspecific DNA-binding interactions occurred between TTHA1437and the ST2R24 template (Figure 12A) or TTHA1719 and the ST2R35 template (Figure 12B). REPA utilizes a fluorescent green-labeled DNA probe, REPSAis, which contains adefined DNA template, the REPSA-selected DNAs (Round 6 or Round 10, respectively),and the respective IISRE. Since no REPSA-selected DNAs were observed for TTHA1437 without cAMP, no REPA was performed. The REPA results for TTHA1437 indicate nonspecific binding due to TTHA1437-REPSAis binding interactions.

Contrarily, the REPA results for TTHA1719 do not indicate any binding interactions at all, specific or nonspecific, since both the cleavage control and experimental reactions do not show any protein-DNA interactions.

In addition to REPA, EMSA was used to validate REPSA selections which would typically quantify the affinity of the TF for the selected DNAs. Ten-fold dilutions of TTHA1437 and TTHA1719 with their respective selected DNAs did now show any protein-DNA complex formation which indicates that the protein is not binding to any or not the majority of the DNA sequences (Figures 13A and 13B).

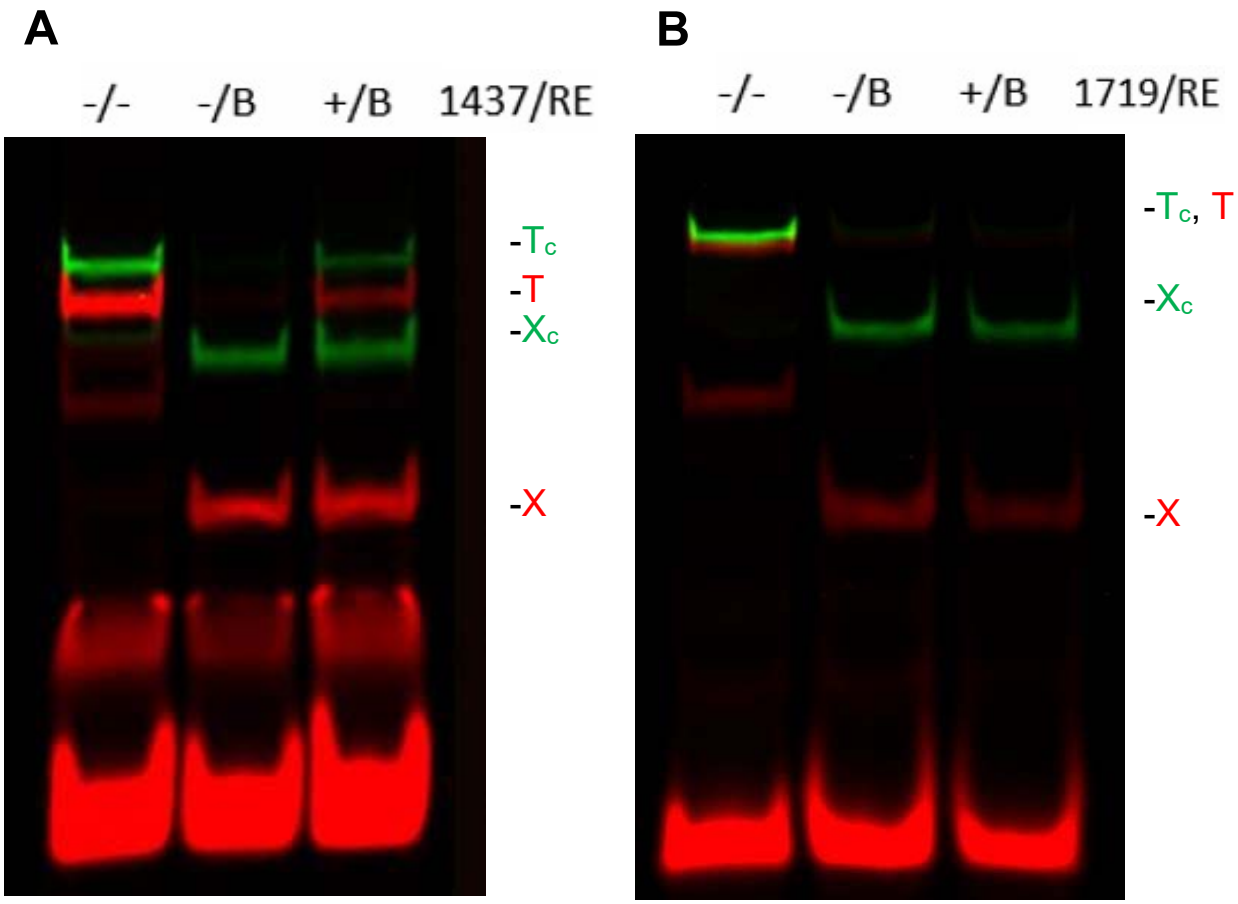

Figure 12. Attempt to validate TTHA1437 and TTHA1719 DNA-binding sequences with REPA. (**A**) Shown is an IR fluorescence image of restriction endonuclease protection assay with ST2R24 DNAs selected from Round 6 of REPSA for TTHA1437 with cAMP. (**B**) Shown is an IR fluorescence image of REPA with ST2R35 DNAs selected from Round 10 of REPSA for TTHA1719. The presence(+) or absence (-) of protein (TTHA1437 or TTHA1719 respectively) and IISRE BpmI (B) are denoted above each lane. Band designations: (T) intact and (X) cleaved IRD7-labeled ST2R24 or ST2R35 selection template (red) and IRD8- labeled REPSAis control DNA (green).

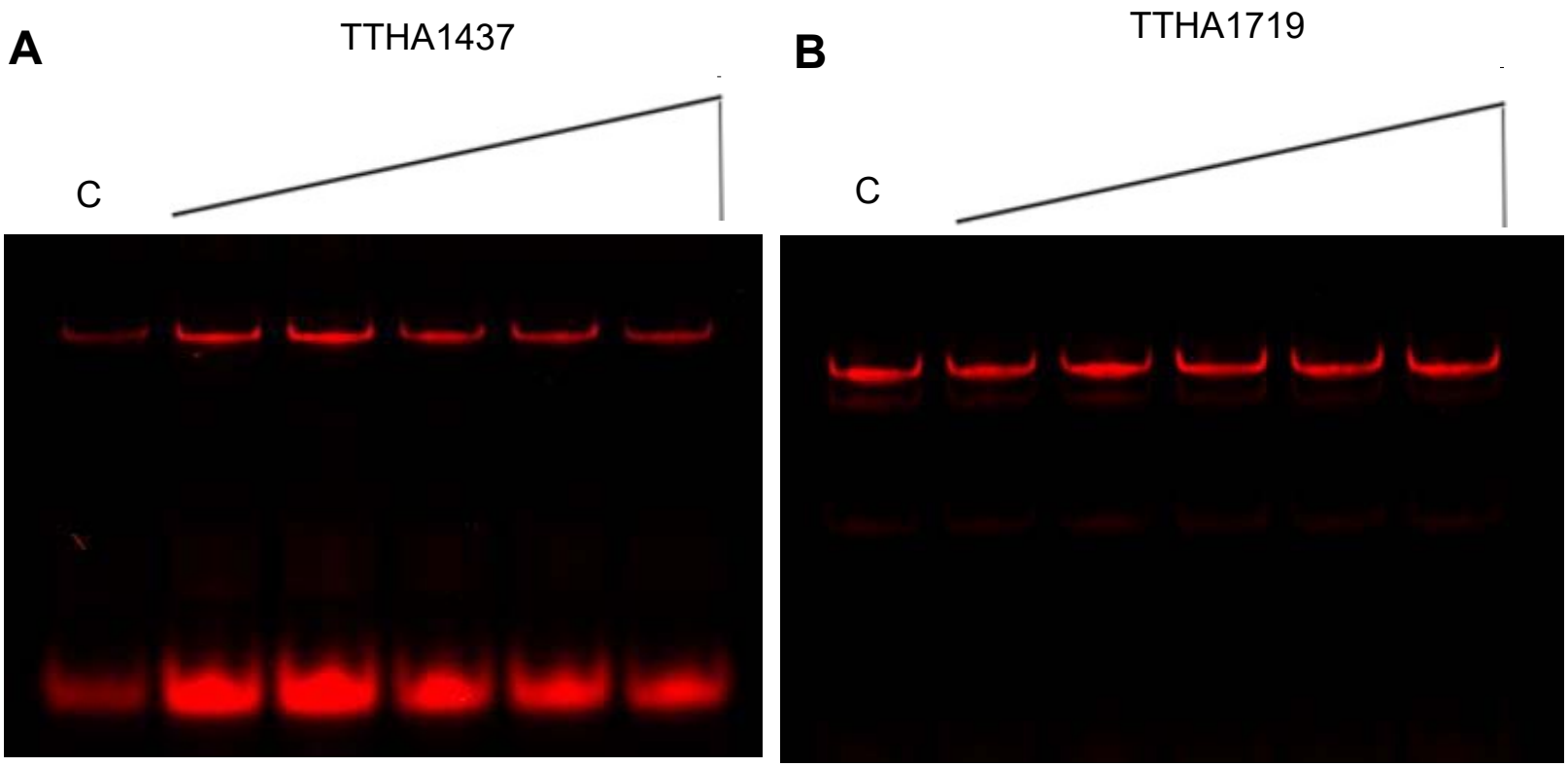

Figure 13. Attempt to validate TTHA1437 and TTHA1719 DNA-binding sequences with EMSA. (**A**) Shown is an IR fluorescence image of electrophoretic mobility shift assay with ST2R24 DNAs selected from Round 6 of REPSA for TTHA1437 with cAMP incubated with increasing concentrations of TTHA1437 protein (from left to right: 0, 3.09, 30.9 309 and 3,090 nM TTHA1437). (**B**) Shown is an IR fluorescence image of EMSA with ST2R35 DNAs selected from Round 10 of REPSA for TTHA1719 incubated with increasing concentrations of TTHA1719 protein (from left to right: 0,4.16, 41.6, 416, and 4,160 nM TTHA1719). (C) indicates the DNA control in the absence of protein.

Chapter 3.4 Analysis of Proposed and Known Consensus DNAs with Binding Assays

Prior to REPSA selections with TTHA1719, the proposed consensus sequence for its binding site was subjected to multiple REPAs to determine the optimal protocol conditions which were later used in REPSA selections.[28] These assays were done prior to REPSA selections due to TTHA1719's low molecular weight and its tetrameric binding activity. The initial REPA was performed under standard and unaltered conditions to create baseline data to compare altered REPA experiments to (Figure 14). The REPA experiments with altered reaction conditions are as follows: addition of a reducing agent (1 mM DTT), increased digestion time for IISRE (7 and 9 min), and altered incubation temperature for TF binding (37°C or 65°C); Figures 15-19. The results of these preliminary experiments demonstrated that the reaction conditions to be used in REPSA selections for TTHA1719 were the addition of 1 mM DTT, 9 min digestion time with IISRE FokI or BpmI, and an incubation temperature of 55°C to allow protein-DNA complex formation.

Additionally, EMSA and BLI were performed for the *E. coli* CRP consensus sequence with TTHA1437 following REPSA selections. These assays were done after REPSA selections and validation since the REPA results showed nonspecific binding and the EMSA results did not show any protein-DNA complex formation (Figure 20). The *E. coli* CRP consensus DNA was chosen for these assays since TTHA1437 belongs to the CRP/FNR class of TFs in *T. thermophilus* HB8 and is speculated to be homologous to the *E. coli* CRP in function and/or structure.[20,22] Unfortunately, the results of EMSA with the CRP consensus sequence and TTHA1437 did not show any protein-DNA complex formation. The BLI results also showed no binding activity when the real-time kinetics

were observed with TTHA1437 concentrations of 11, 33, 100 and 300 nM (Figure 21). No BLI was done with our control DNA, REPSAis, since there BLI results between TTHA1437 and the *E. coli* CRP consensus DNA were ambiguous.

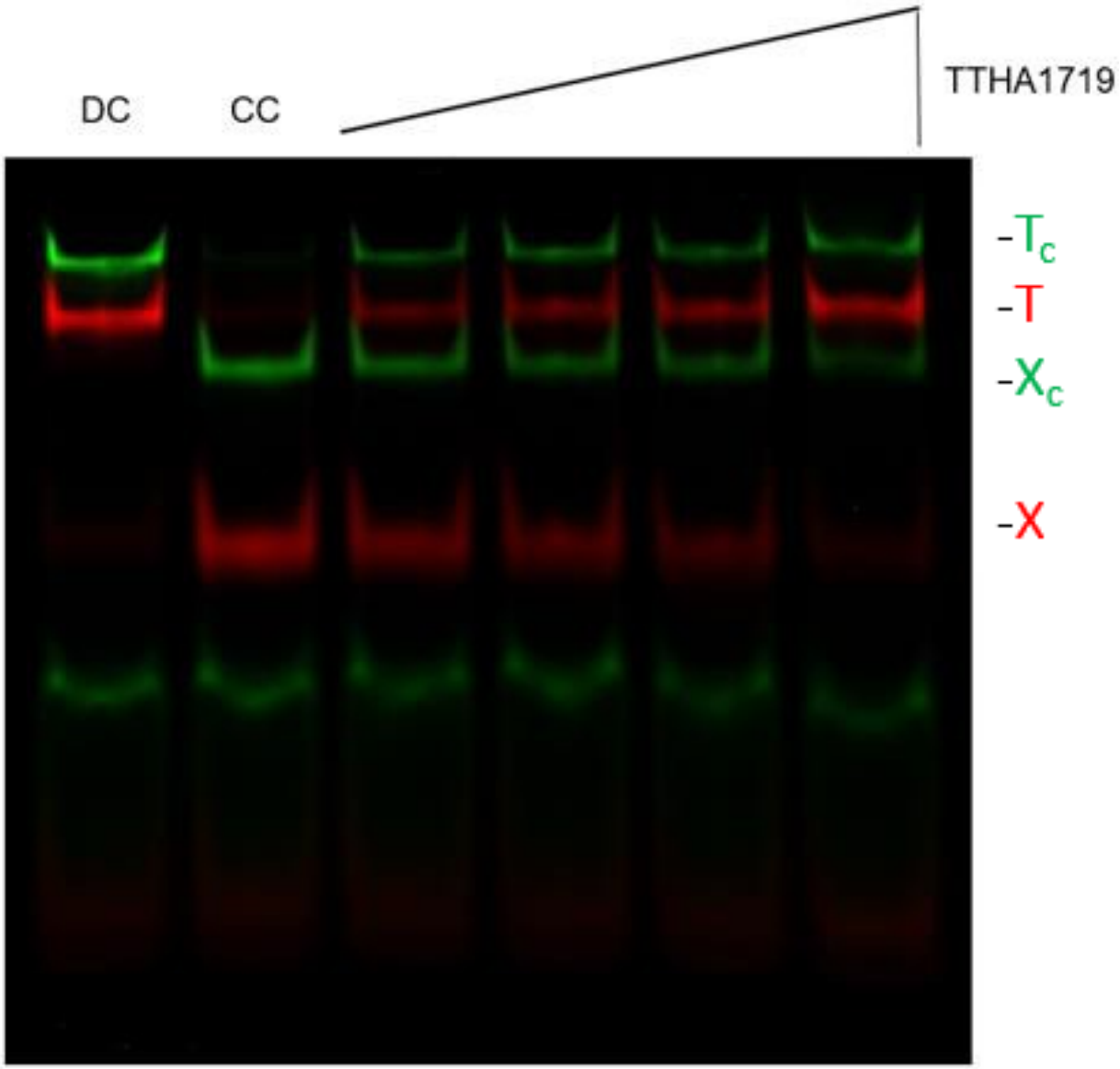

Figure 14. Baseline REPA with proposed consensus and TTHA1719. Shown is an IR fluorescence image of restriction endonuclease protection assay with ST2_1719_jcon-60 and TTHA1719. (DC) denotes DNA control, (CC) denotes the cleavage control. Lanes 2-6 contain 0.4 U IISRE FokI. TTHA1719 concentrations from left to right are 4.16, 41.6, 416, and 4160 nM. Band designations: (T) intact and (X) cleaved IRD7-labeled ST2_1719_jcon-60 DNA template (red) and IRD8-labeled REPSAis control DNA (green).

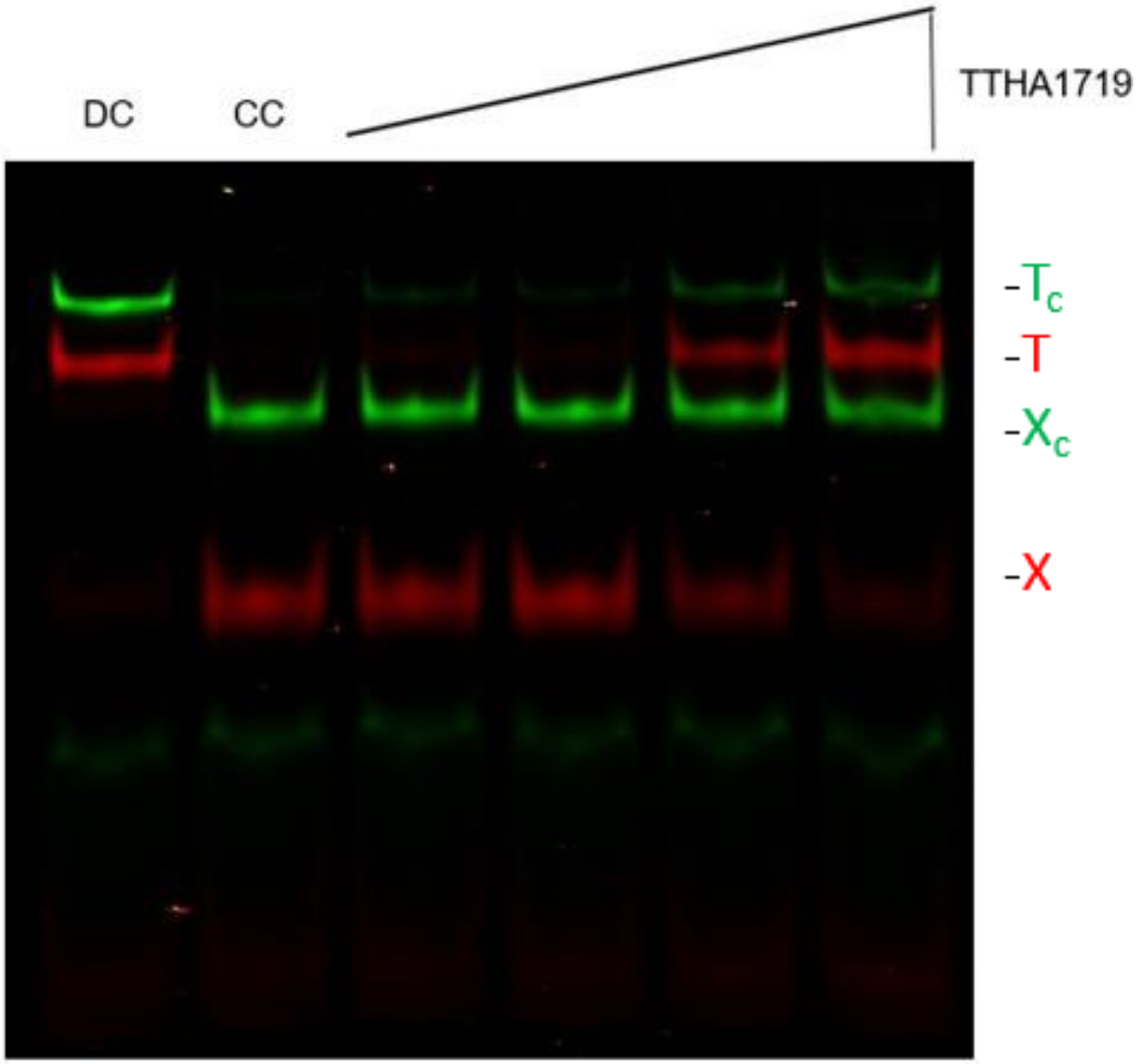

Figure 15. Experimental REPA with a reducing agent. Shown is an IR fluorescence image of restriction endonuclease protection assay with ST2_1719_jcon-60, TTHA1719, and 1 mM DTT. (DC) denotes DNA control, (CC) denotes the cleavage control. Lanes 2-6 contain 0.4 U IISRE FokI. TTHA1719 concentrations from left to right are 4.16, 41.6, 416, and 4160 nM. Band designations: (T) intact and (X) cleaved IRD7-labeled ST2_1719_jcon-60 DNA template (red) and IRD8-labeled REPSAis control DNA (green).

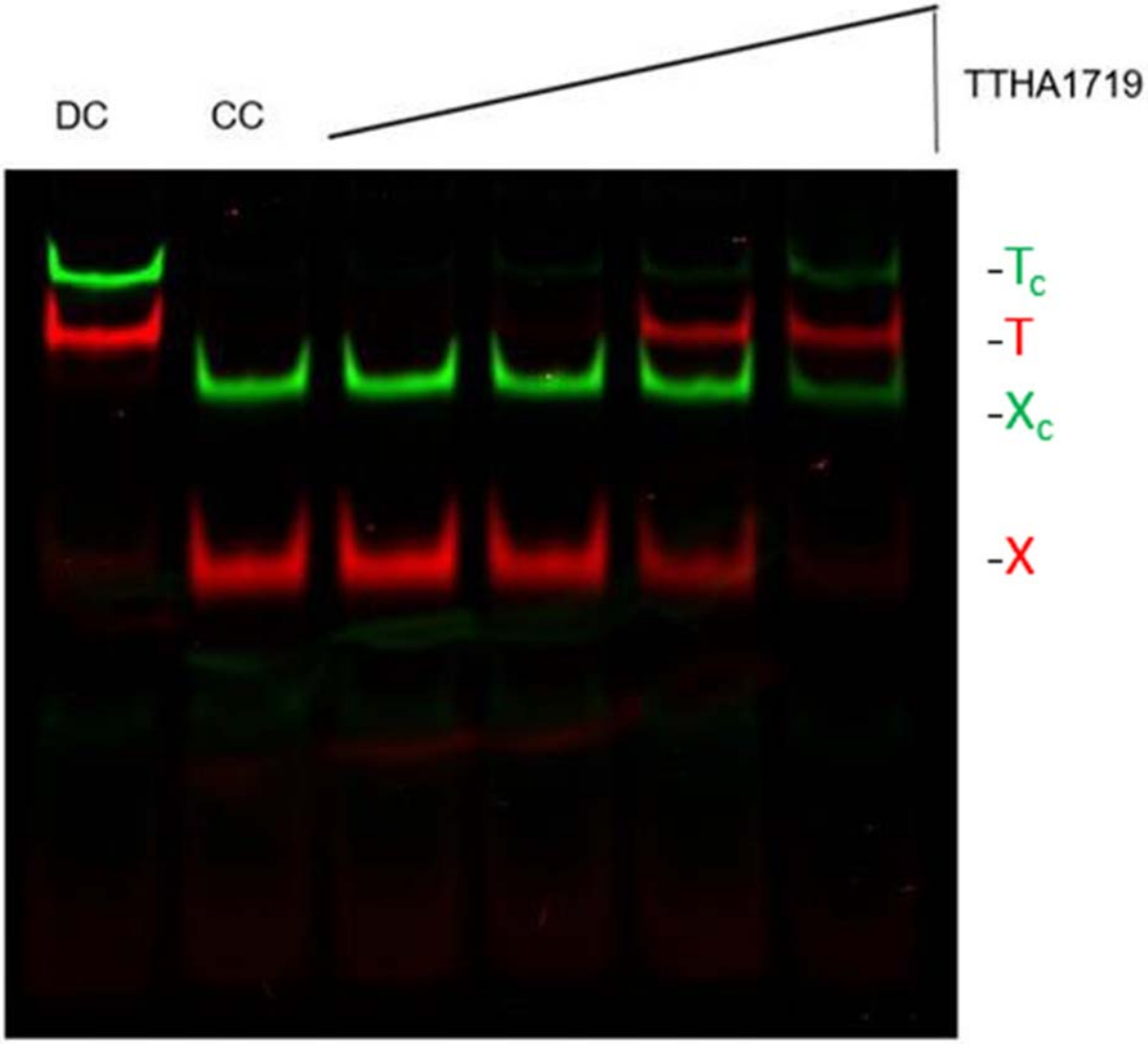

Figure 16. Experimental REPA with 7 min IISRE digestion time. Shown is an IR fluorescence image of restriction endonuclease protection assay with ST2_1719_jcon-60, TTHA1719, and 1 mM DTT. (DC) denotes DNA control, (CC) denotes the cleavage control. Lanes 2-6 contain 0.4 U IISRE FokI. TTHA1719 concentrations from left to right are 4.16, 41.6, 416, and 4160 nM. Band designations: (T) intact and (X) cleaved IRD7-labeled ST2_1719_jcon-60 DNA template (red) and IRD8-labeled REPSAis control DNA (green).

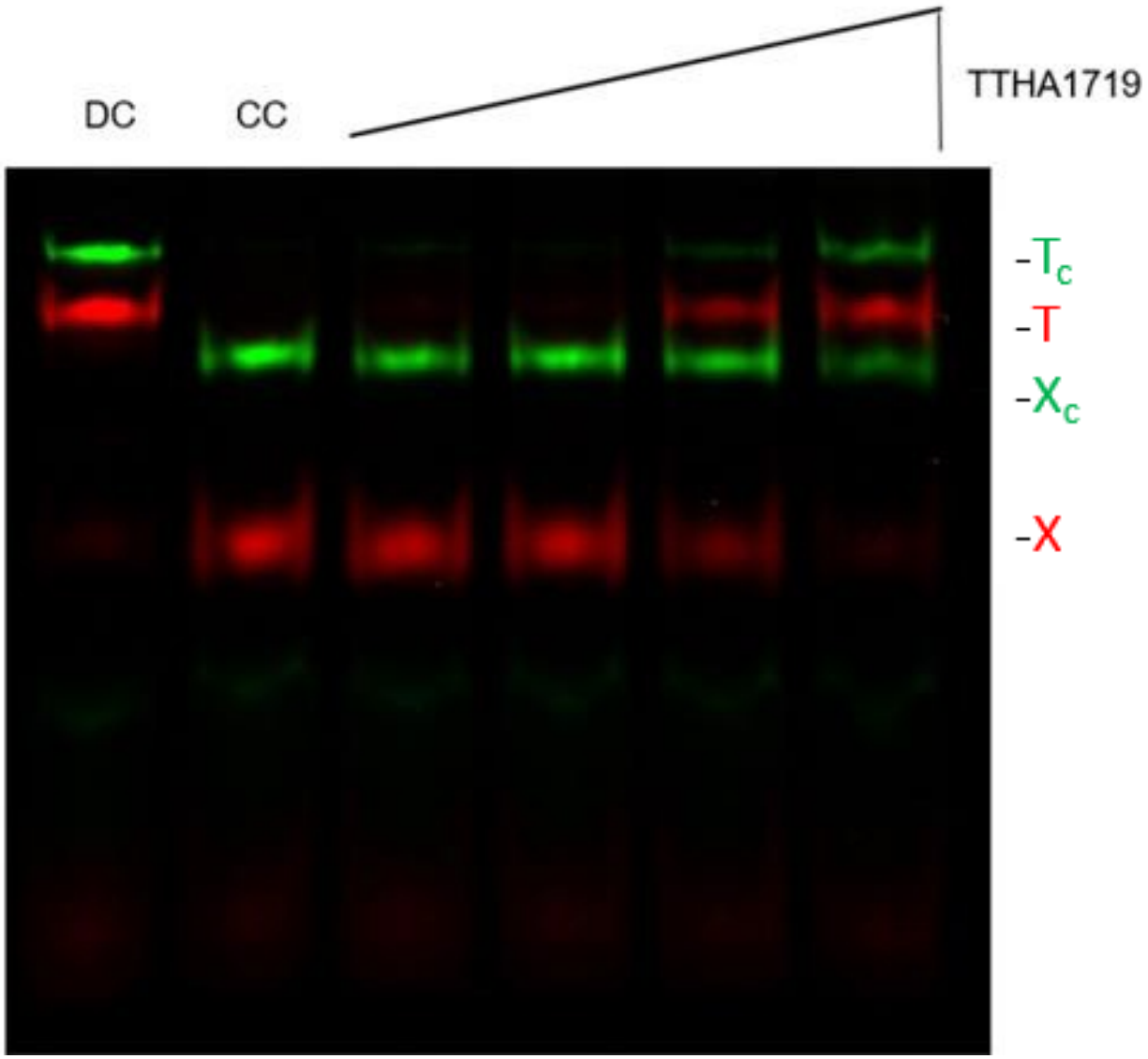

Figure 17. Experimental REPA with 9 min IISRE digestion time. Shown is an IR fluorescence image of restriction endonuclease protection assay with ST2_1719_jcon-60, TTHA1719, and 1 mM DTT. (DC) denotes DNA control, (CC) denotes the cleavage control. Lanes 2-6 contain 0.4 U IISRE FokI. TTHA1719 concentrations from left to right are 4.16, 41.6, 416, and 4160 nM. Band designations: (T) intact and (X) cleaved IRD7-labeled ST2_1719_jcon-60 DNA template (red) and IRD8-labeled REPSAis control DNA (green).

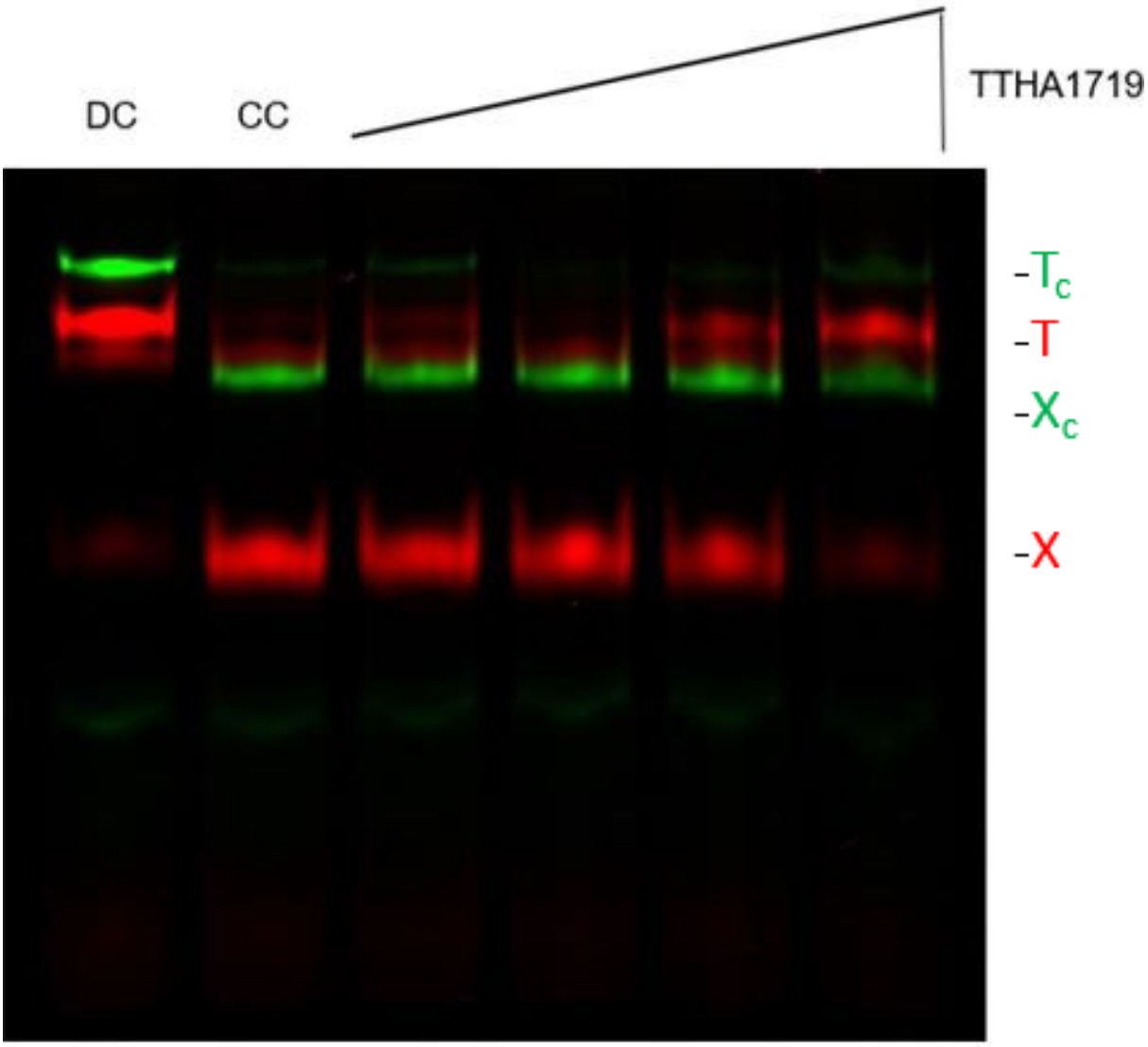

Figure 18. Experimental REPA with 37°C binding temperature. Shown is an IR fluorescence image of restriction endonuclease protection assay with ST2_1719_jcon-60, TTHA1719, and 1 mM DTT. (DC) denotes DNA control, (CC) denotes the cleavage control. Lanes 2-6 contain 0.4 U IISRE FokI. TTHA1719 concentrations from left to right are 4.16, 41.6, 416, and 4160 nM. Band designations: (T) intact and (X) cleaved IRD7-labeled ST2_1719_jcon-60 DNA template (red) and IRD8-labeled REPSAis control DNA (green).

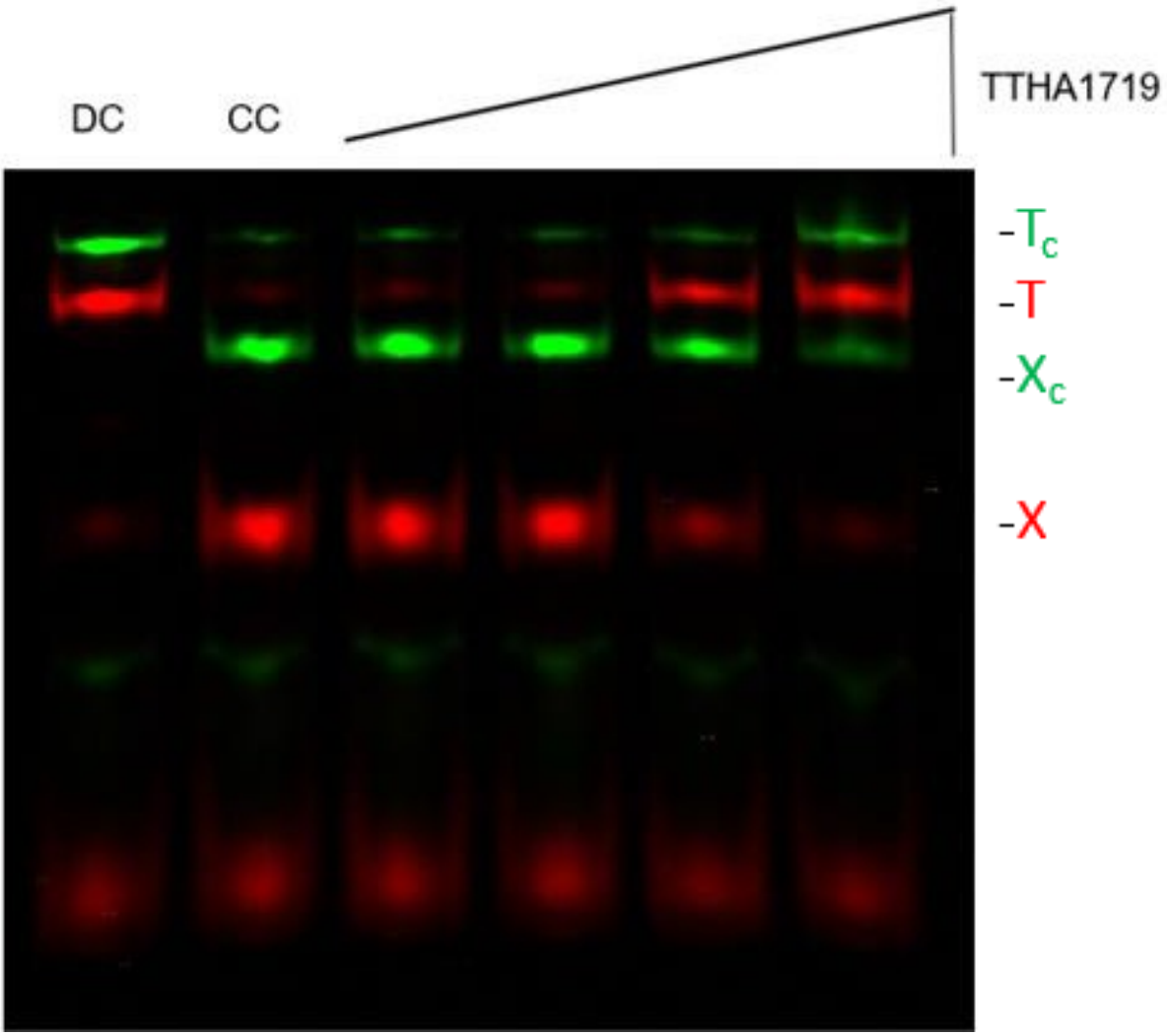

Figure 19. Experimental REPA with 65°C binding temperature. Shown is an IR fluorescence image of restriction endonuclease protection assay with ST2_1719_jcon-60, TTHA1719, and 1 mM DTT. (DC) denotes DNA control, (CC) denotes the cleavage control. Lanes 2-6 contain 0.4 U IISRE FokI. TTHA1719 concentrations from left to right are 4.16, 41.6, 416, and 4160 nM. Band designations: (T) intact and (X) cleaved IRD7-labeled ST2_1719_jcon-60 DNA template (red) and IRD8-labeled REPSAis control DNA (green).

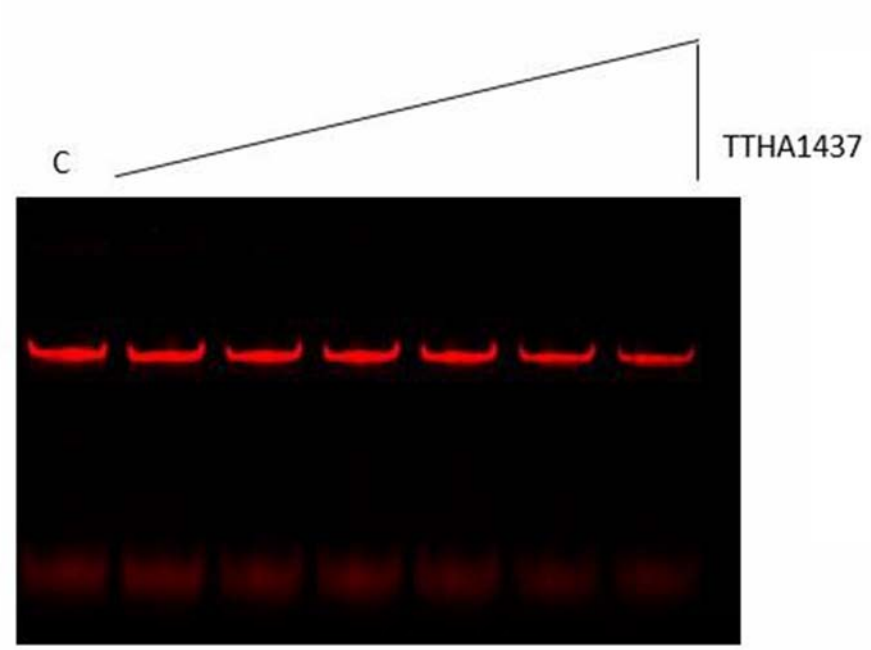

Figure 20. Investigation of TTHA1437-CRP consensus interactions with EMSA. Shown is an IR fluorescence image of electrophoretic mobility shift assay with IRD7-labeled ST2_CRP_Ec DNA and TTHA1437. Thereactions were run with 2 mM cAMP and the running buffer was supplemented with cAMP to a final concentration of 2 mM. No protein-DNA complex formation was observed.

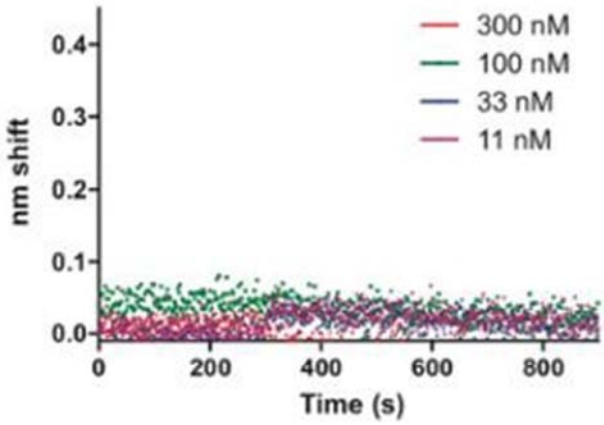

Figure 21. BLI raw data plot of TTHA1719-CRP consensus binding. Shown is the biolayer interferometry data between TTHA1719 and the *E. coli* consensus DNA. Binding kinetics were observed over 1200 sec with TTHA1437 concentrations of 11, 33, 100, and 300 nM. The BLI 100 buffer was supplemented with cAMP to a final concentration of 2 mM. Since no association or dissociation steps were observed, no K_D could be calculated and the results of this BLI were deemed ambiguous.

DISCUSSION

In these studies, the combinatorial selection method Restriction Endonuclease Protection, Selection, and Amplification (REPSA), was used to determine if DNA-binding interactions occurred between our ST2R24 selection template and TTHA1417 as well as the ST2R35 selection template and TTHA1719. Although TTHA1417 showed a ~30% protected DNA pool and TTHA1719 showed ~10% protected DNA pool when the results were verified using Restriction Endonuclease Protection Assay (REPA) and Electrophoretic Mobility Shift Assay (EMSA) the selected species were either deemed as nonspecific binding or not able to be validated using these methods. As mentioned previously, using EMSA to validate TTHA1437-DNA interactions was not a preferred method due to its proposed need for the effector molecule cAMP.

The preliminary REPA binding assays with the proposed consensus DNA for TTHA1719 proved to be promising with the amount of protected DNA species, but this was only used to establish optimal protocol conditions for REPSA selections with TTHA1719. If time allowed, it would have been beneficial to run other binding assays such as EMSA to determine the percentage of DNA-protein complexes as well as Biolayer Interferometry (BLI) to ascertain the binding kinetics.

Although our REPSA method has worked for previously studied *Thermus thermophilus* HB8 transcription factors in our laboratory, REPSA was not successful in identifying the DNA-binding consensus sequence for both TTHA1417 and TTHA1719. This demonstrates that this method is not viable for identifying and characterizing all TFs

in *T. thermophilus* HB8. As previously stated, TTHA1417 belongs to the CRP/FNR family of activators in this organism and has been proposed to be a cAMP receptor protein with possible homology to *E. coli* CRP. It is possible that TTHA1437 is not a local regulator, but a global regulator like that of *E. coli* CRP. If this is the case, the REPSA method would not be able to identify a consensus DNA-binding sequence for TTHA1437 due to many factors such as co-regulation, dual regulation and a possible DNA-binding domain that is not highly specific.

TTHA1719 is a copper-sensing transcriptional repressor (CsoR) which is speculated to only regulate a single operon, the copper-sensitive operon. Using REPAS to identify a TF which regulates only a single operon would be extremely difficult since its DNA-binding sequence would be extremely specific to that transcription factor. In the past, TFs characterized by REPSA have traditionally been thought to regulate more than a single operon, with most potentially regulating multiple. Although REPSA was unable to identify and characterize the DNA-binding sites of TTHA1437 and TTHA1719 and their potential biological functions, these studies show that the REPSA method can in fact not characterize all *Thermus thermophilus* HB8 transcription factors.

APPENDIX: FUTURE DIRECTIONS

Future directions include further investigation of the TTHA1719 proposed consensus sequence as well as the investigation of the asterisk species found in REPSA selections with both ST2R24 and TTHA1437 as well as ST2R35 and TTHA1719. The asterisk species has rarely been observed and its sequence is highly unknown. It is a truncated DNA species that is caused by internal priming in the randomized region of the selection template. Interestingly, the asterisk species became visible and amplified more rapidly with the ST2R35 template than the ST2R24 template, so it is quite possible that the length of the internal randomized region plays a role in asterisk species formation. Sequencing these rare DNA species would determine if the asterisk species from different REPSA selections and selection templates would reveal different sequences.

REFERENCES

1. Venter, J. C.; Smith, H. O.; Adams, M. D. The Sequence of the Human Genome. *Clin. Chem.* **2015***, 61*, 1207-1208.

2. Babu, M. M.; Balaji, S.; Aravind, L. General trends in the evolution of prokaryotic transcriptional regulatory networks. *Genome Dyn.* **2007***, 3*, 66-80.

3. Sahota, G.; Stormo, G. D. Novel sequence-based method for identifying transcription factor binding sites in prokaryotic genomes. *Bioinformatics* **2010***, 26*, 2672-2677.

4. Dey, B.; Thukral, S.; Krishnan, S.; Chakrobarty, M.; Gupta, S.; Manghani, C.; Rani, V. DNA-protein interactions: methods for detection and analysis. *Mol. Cell. Biochem.* **2012***, 365*, 279-299.

5. Charoensawan, V.; Wilson, D.; Teichmann, S. A. Lineage-specific expansion of DNA-binding transcription factor families. *Trends Genet.* **2010***, 26*, 388-393.

6. Hung, J. H.; Weng, Z. Discovering cis-Regulatory Motifs. *Cold Spring Harb Protoc.* **2017***, 2017*, 10.1101/pdb.prot093203.

7. Agari, Y.; Kashihara, A.; Yokoyama, S.; Kuramitsu, S.; Shinkai, A. Global gene expression mediated by Thermus thermophilus SdrP, a CRP/FNR family transcriptional regulator. *Mol. Microbiol.* **2008***, 70*, 60-75.

8. Moses, R. E.; O'Malley, B. W. DNA transcription and repair: a confluence. *J. Biol. Chem.* **2012***, 287*, 23266-23270.

9. Schleif, R. F. Modulation of DNA binding by gene-specific transcription factors. *Biochemistry* **2013***, 52*, 6755-6765.

10. Borukhov, S.; Nudler, E. RNA polymerase holoenzyme: structure, function and biological implications. *Curr. Opin. Microbiol.* **2003***, 6*, 93-100.

11. Xue, Y.; Hogan, B. P.; Erie, D. A. Purification and initial characterization of RNA polymerase from Thermus thermophilus strain HB8. *Biochemistry* **2000***, 39*, 14356-14362.

12. Beckwith, J. R. Regulation of the lac operon. Recent studies on the regulation of lactose metabolism in Escherichia coli support the operon model. *Science* **1967***, 156*, 597-604.

13. Meiklejohn, A. L.; Gralla, J. D. Activation of the lac promoter and its variants. Synergistic effects of catabolite activator protein and supercoiling in vitro. *J. Mol. Biol.* **1989***, 207*, 661-673.

14. Ishihama, A. Building a complete image of genome regulation in the model organism Escherichia coli. *J. Gen. Appl. Microbiol.* **2018***, 63*, 311-324.

15. Lewis, M. Allostery and the lac Operon. *J. Mol. Biol.* **2013***, 425*, 2309-2316.

16. Santos, C. L.; Tavares, F.; Thioulouse, J.; Normand, P. A phylogenomic analysis of bacterial helix-turn-helix transcription factors. *FEMS Microbiol. Rev.* **2009***, 33*, 411-429.

17. Huffman, J. L.; Brennan, R. G. Prokaryotic transcription regulators: more than just the helix-turn-helix motif. *Curr. Opin. Struct. Biol.* **2002***, 12*, 98-106.

18. Suvorova, I. A.; Korostelev, Y. D.; Gelfand, M. S. GntR Family of Bacterial Transcription Factors and Their DNA Binding Motifs: Structure, Positioning and Co-Evolution. *PLoS One* **2015***, 10*, e0132618.

19. Beckwith, J. A genetic approach to characterizing complex promoters in E. coli. *Cell*

1981*, 23*, 307-308.

20. Crystal structure of the E. coli CRP-cAMP complex. https://www.rcsb.org/structure/2GZW

21. Basak, S.; Jiang, R. Enhancing E. coli tolerance towards oxidative stress via engineering its global regulator cAMP receptor protein (CRP). *PLoS One* **2012***, 7*, e51179.

22. Zheng, D.; Constantinidou, C.; Hobman, J. L.; Minchin, S. D. Identification of the CRP regulon using in vitro and in vivo transcriptional profiling. *Nucleic Acids Res.* **2004***, 32*, 5874-5893.

23. Cava, F.; Hidalgo, A.; Berenguer, J. Thermus thermophilus as biological model. *Extremophiles* **2009***, 13*, 213-231.

24. Ohtani, N.; Tomita, M.; Itaya, M. An extreme thermophile, Thermus thermophilus, is a polyploid bacterium. *J. Bacteriol.* **2010***, 192*, 5499-5505.

25. Oshima, T.; Imahori, K. Description of Thermus thermophilus (Yoshida and Oshima) comb. nov., a nonsporulating thermophilic bacterium from a Japanese thermal spa. *Int. J. Syst. Evol. Microbiol*. **1974**, 24, 102-112.

26. Shinkai, A.; Kira, S.; Nakagawa, N.; Kashihara, A.; Kuramitsu, S.; Yokoyama, S. Transcription activation mediated by a cyclic AMP receptor protein from Thermus thermophilus HB8. *J. Bacteriol.* **2007***, 189*, 3891-3901.

27. Green, J.; Stapleton, M. R.; Smith, L. J.; Artymiuk, P. J.; Kahramanoglou, C.; Hunt, D. M.; Buxton, R. S. Cyclic-AMP and bacterial cyclic-AMP receptor proteins revisited: adaptation for different ecological niches. *Curr. Opin. Microbiol.* **2014***, 18*, 1-7.

28. Sakamoto, K.; Agari, Y.; Agari, K.; Kuramitsu, S.; Shinkai, A. Structural and functional characterization of the transcriptional repressor CsoR from Thermus thermophilus HB8. *Microbiology (Reading)* **2010***, 156*, 1993-2005.

29. Van Dyke, M. W.; Van Dyke, N.; Sunavala-Dossabhoy, G. REPSA: General combinatorial approach for identifying preferred ligand–DNA binding sequences. *Methods.* **2007***, 42*, 118-127.

30. Szybalski, W.; Kim, S. C.; Hasan, N.; Podhajska, A. J. Class-IIS restriction enzymes--a review. *Gene* **1991***, 100*, 13-26.

31. Hardenbol, P.; Van Dyke, M. W. Sequence specificity of triplex DNA formation: Analysis by a combinatorial approach, restriction endonuclease protection selection and amplification. *Proc. Natl. Acad. Sci. U. S. A.* **1996***, 93*, 2811-2816.

32. Van Dyke, M. W.; Beyer, M. D.; Clay, E.; Hiam, K. J.; McMurry, J. L.; Xie, Y. Identification of Preferred DNA-Binding Sites for the Thermus thermophilus Transcriptional Regulator SbtR by the Combinatorial Approach REPSA. *PLoS One* **2016***, 11*, e0159408.

33. Lee, M.; Um, H.; Van Dyke, M. W. Identification and characterization of preferred DNA-binding sites for the Thermus thermophilus transcriptional regulator FadR. *PLoS One* **2017***, 12*, e0184796.

34. Hellman, L. M.; Fried, M. G. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat. Protoc.* **2007***, 2*, 1849-1861.

35. Cox, J. S.; Moncja, K.; Mckinnes, M.; Van Dyke, M. W. Identification and

Characterization of Preferred DNA-Binding Sites for the Thermus thermophilus HB8 Transcriptional Regulator TTHA0973. *Int. J. Mol. Sci.* **2019***, 20*, 10.3390/ijms20133336.

36. Moncja, K.; Van Dyke, M. W. Determination and Dissection of DNA-Binding Specificity for the Thermus thermophilus HB8 Transcriptional Regulator TTHB099. *Int. J. Mol. Sci.* **2020***, 21*, 10.3390/ijms21217929.